Technische Universität München

Max-Planck-Institut für Biochemie

Abteilung Strukturforschung

Biologische NMR-Arbeitsgruppe

Biochemical and biophysical characterization

of the lyase isomerase PecE/PecF complex, nicastrin – the

transmembrane component of the γ-secretase complex

and

structural investigations of the genomic islands' integrases

Aleksandra Szwagierczak

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.–Prof. Dr. M. Groll

Prüfer der Dissertation: 1. Priv.-Doz. Dr. N. Budisa

2. Univ.–Prof. Dr. Chr.F.W. Becker

Die Dissertation wurde am 28.09.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie a m 04.11.2009 angenommen.

Acknowledgements

I would like to express my gratitude to everyone who contributed to this work.

In particular, I am grateful to Professor Nediljko Budisa for being my Doktorvater.

To my supervisor Doctor Tad A. Holak for his support, discussions and interest in my work.

To the laboratory team: Michał Biśta, Kinga Brongel, Anna Czarny, Anna Ducka, Weronika Janczyk, Kaja Kowalska, Aleksandra Mikołajka, Marcin Krajewski, Grzegorz Popowicz, Ulli Rothweiler, Arkadiusz Sikora, Tomasz Sitar, Siglinde Wolf for their help, advice, and good working atmosphere.

I would especially like to thank Kinga and Igor for their friendship and support.

Last but not least, I would like to acknowledge my family for all the care, support, and encouragement.

Publications

Parts of this thesis have been or will be published in due course:

Grzegorz M. Popowicz, Anna Czarna, Ulli Rothweiler, **Aleksandra Szwagierczak**, Marcin Krajewski, Lutz Weber, Tad A. Holak (2007) Molecular basis for the inhibition of p53 by Mdmx. *Cell Cycle* **6**, 2386-2392.

Aleksandra Szwagierczak, Uladzimir Antonenka, Grzegorz M. Popowicz, Tomasz Sitar, Tad A. Holak, Alexander Rakin (2009) Structures of the arm-type binding domains of HPI and HAI7 integrases. *J. Biol. Chem.* **284**, 31664-31671.

Aleksandra Szwagierczak, Uladzimir Antonenka, Grzegorz M. Popowicz, Alexander Rakin, Tad A. Holak, X-ray structure of arm-type binding domain of HPI integrase in complex with DNA. In preparation.

Contents

1 Introduction

1.1 Nicastrin, the component of γ-secretase

1.1.1 Intramembrane proteolysis

Classical proteases, having necessity for aqueous environment, carry out catalysis in cytoplasmic, lumenal, or extracellular environment. Recently, however, a novel group of polytopic membrane proteases was identified which have an active site buried within the membrane. These enzymes are able to cleave a peptide bond located within the lipid bilayer and thus were designated the intramembrane-cleaving proteases (I-CLiPs). Consistent with use of conserved catalytic mechanisms, structural studies suggest that the I-CLiPs form a pore-like structure within the membrane, permitting water access to their active site.

Several families of I-CLIPs were identified so far: the site 2 protease (S2P) family of zinc metalloproteases, the Rhomboid family of serine proteases, the signal peptide peptidases (SPPs) and the presenilins (both being aspartyl proteases). Regulated intramembrane proteolysis (RIP) performed by these enzymes functions in various processes ranging from bacteria to man. For example, the zinc protease S2P cleaves the sterol-regulatory-element binding protein (SREBP) and ATF6, which results in upregulation of genes involved in cholesterol biosynthesis and the unfolded protein response, respectively (Brown and Goldstein, 1999, Liu and Kaufman, 2003). SPPs catalyze the proteolysis of remnant signal peptides after they have been cleaved from their precursor by signal peptidases (Xia and Wolfe, 2003). The rhomboid family proteases participate in diverse biological pathways such as host cell

invasion and quorum-sensing in bacteria and signaling via the epidermal growth factor receptor in *Drosophila melanogaster* (Freeman, 2008).

In many recognized substrates for S2P, Rhomboid, and SPP, helix-breaking residues are critical determinants for substrate specificity (Beel et al., 2008). This, however, does not appear to be the case for γ-secretase substrates, which are single-pass transmembrane proteins that have shed their extracellular domains through metalloprotease-mediated cleavage close to the membrane, leaving transmembrane–intracellular domain fragments with sequences having little in common. Therefore, with such a relaxed substrate specificity γ-secretase was proposed to function as a 'membrane proteasome' to clear the membrane of processed transmembrane proteins. Nevertheless, in several cases, it is known that the intramembrane cleavage results in the release of an intracellular domain (ICD), which is involved in nuclear signaling and transcriptional regulation. Two most notable γ–secretase substrates are the Alzheimer precursor protein and signaling-involved Notch protein. The former gives rise to neurotoxic β-amyloid peptide, while the cleavage of the latter liberates ICD that serves as transcription factor.

1.1.2 γ-secretase composition, structure, and activity

γ-secretase minimal active composition consists of four proteins: presenilin (PS), nicastrin (NCT), the anterior pharynx defective–1 (Aph-1), and the presenilin enhancer–2 (PEN-2). Several other proteins were shown to interact with γ-secretase, e.g.: TM21 and CD147 (both negative regulators of γ-secretase) or Rer1, which helps to retain the unassembled PEN-2 and NCT within the ER.

As expected from components of a membrane-embedded enzyme, all its components contain transmembrane domains (TMs). Presenilin, identified as an

enzymatic activity-bearing component contains nine of them. It is present in the complex in a heterodimeric form produced by cleavage in the large cytosolic loop connecting TMs 6 and 7. Another distinctive feature of PS is a hydrophobic domain placed in the same loop (HDVII).

Aph-1 is a seven TM-domain protein with the N-terminus located in the lumen and the C-terminus in the cytosol, while Pen-2 is a small hairpin membrane protein with both the N- and C-termini facing the extracellular space.

NCT is characterized as a type I transmembrane protein (i.e. spanning the plasma membrane only once, with the N-terminus facing the lumen) with a bulky, glycosylated ectodomain.

Presenilin and Aph-1 in humans are represented by two homologs each (named PS1 and PS2 and Aph-1a and Aph-1b, respectively). Additionally, Aph-1a exists in long (Aph-1aL) and short (Aph-1aS) splice variants differing in the length of the C-terminus. Because neither the catalytic subunits PS1 and PS2 nor Aph-1a and Aph-1b occur in the same complexes, a minimal set of at least six distinct γ-secretase complexes exist in human cells.

1.1.2.1 Active site and docking site

The active site of γ-secretase was identified as presenilin's highly conserved Asp237 and Asp385, making it an aspartic protease. The former of these aspartates is embedded in a conserved YD motif, whereas the latter is part of the equally conserved GxGD motif. Catalytic residues Asp237 and Asp385 are positioned very closely and opposing each other (at least in the active conformation) on TM6 and 7, respectively. These TMs, along with TM9, were shown to contribute to a water-containing, funnel-shaped cavity inside the membrane (Sato et al. 2006, Tolia et al., 2008). HDVII is embedded in the cavity in such a way that it conceals TM7 from

interacting with TM9 or TM8. The TM9's highly conservative motif, PAL, contains a proline residue which is known to induce flexible kinks in transmembrane helices. Thanks to this flexibility TM9 is moving upon the substrate binding, bringing the substrate inside the cavity for the cleavage to occur (Tolia et al., 2008).

Apart from the protease active site, γ-secretase possesses a distinct substrate-binding site. This docking site is localized on the interface between NTF and CTF of presenilin and is in close proximity (distance equal or smaller to 3 residues of a polypeptide) to the active site (Kornilova et al., 2005). The GxGD motif was proposed to be part of the docking site with the residue x being critical for the APP/Notch substrate selectivity (Yamasaki et al., 2006).

1.1.2.2 Structure

The low-resolution 3D-EM reveals the shape of γ-secretase as an elongated globule with the longest dimension of 120 Å (in the membrane) and the other two dimensions, perpendicular to the long axis, being 70–80 Å each. A belt-like density 60 Å high, visible around the particle is, most likely, the membrane-embedded portion. The structure reveals also the presence of a low-density central chamber of an irregular shape and diameter of 20-40 Å. Two pores, which are probably involved in release of proteolytic fragments, lead to the cavity, the apical pore having a diameter of 20 Å, while basal one of 10 Å.

The 3D-EM structure shows no apparent opening in the transmembrane portion of the complex through which a docked TMD substrate could translocate into the central chamber, however two thin density regions identified in the transmembrane portion of the complex may serve as a lateral gates for the substrate to reach the catalytic site. Extracellular density was identified as NCT and it was suggested that it serves as a type of a flexible lid that could regulate the entry of

water molecules into the central chamber and the exit of the hydrophilic ectodomain products (Lazarov et al., 2006).

1.1.2.3 The assembly of γ-secretase

Assembly of γ-secretase is most likely initiated through the formation of a stable Aph-1 and nicastrin subcomplex, followed by entry of the presenilin holoprotein to form a trimeric intermediate complex. The nicastrin : Aph-1 subcomplex contains the immature, partially N-glycosylated nicastrin and might represent an ER-localized step of complex formation, with the subsequent addition of presenilin promoting translocation of the maturing complex to the Golgi, *trans*-Golgi, and the plasma membrane. Pen-2 acts primarily at later steps and is incorporated into intermediate complex. It provokes the endoproteolysis of the presenilin holoprotein generating a 30 kDa N-terminal fragment (NTF) and a 20 kDa C-terminal fragment (CTF) conferring proteolytic activity (Periz and Fortini, 2004).
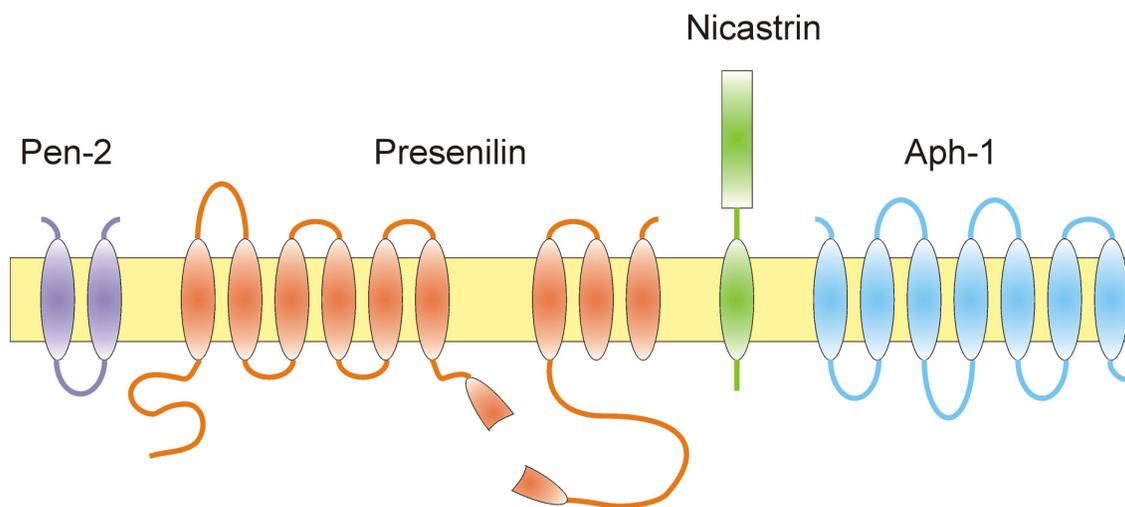
Figure 1.1.1. Schematic representation of γ-secretase components

1.1.3 Intermolecular interactions

Due to the lack of a high-resolution 3D structure of γ-secretase or any of its components, the spatial organization and interaction between proteins can only be characterized through mutation and cross-linking experiments.

1.1.3.1 Presenilin interactions

The N-terminal two-thirds of the TMD4 of PS1 is the direct binding site of Pen-2, with conserved motifs NDTxN and WNF necessary for the interaction (Watanabe et al., 2005).

The C-terminal tail (last 8 residues), together with TM1 and parts of following loop, has been shown to bind APP, which implies a ring structure topology for presenilin (Annaert et al., 2001). Last 6 residues (FHQFY) bind the transmembrane helix of NCT, which implies that these residues dive into or penetrate the membrane. Furthermore, 20 residues positioned on TM9 and beyond it were identified as the ER retention signal that prevents unassembled subunits from getting exported out of the ER and is masked upon the full complex assembly (Kaefer et al. 2004).

1.1.3.2 Aph-1 interactions

Aph-1 does not hold catalytic activities and, as the most stable component of the complex, it might serve as an initial scaffold for its assembly. A GxxxG motif, localized on TM4 of Aph-1, is crucial for the interaction with presenilin and PEN-2, and therefore is necessary for the assembly and activity of γ-secretase; however, not for the NCT binding. This motif can be also responsible for the interaction with the other TMs bearing (small aminoacid)xxx(small aminoacid) motif (for example TM1, 2 and 3 of APH-1) (Niimura et al., 2005). The residue His171 on TM5 was identified as significant in either the Aph-1 trafficking or the interactions with regulatory proteins

controlling the Aph-1 movement in the cell since, if mutated, it caused Aph-1 to interact exclusively with the already matured, cell membrane-associated NCT. His197 from TM6, on the other hand, is suspected to participate in the hydrophilic presenilin cavity (Pardossi et al., 2009b). Regions of Aph-1 interacting with NCT are, so far, unknown.

1.1.3.3 PEN-2 interactions

The extracellular N-terminus PEN-2 has been implicated in the interaction with presenilin (Crystal et al., 2003), while the proximal two-thirds of TM1 are necessary for the endoproteolysis of presenilin. The C-terminus of PEN-2 seems to be responsible for stability of cleaved derivatives of presenilin (Kim et al., 2005).

1.1.4 Nicastrin

Nicastrin was discovered biochemically, by affinity chromatography, as a major presenilin-interacting protein (Yu et al., 2000). The gene located on chromosome 1 codes a 709 amino acid long protein. NCT consists of a large, extracellularly located ectodomain, which is followed by a 20 aminoacids-long hydrophobic transmembrane helix and a short, cytoplasmic C-terminus of 20 residues.

1.1.4.1 Glycosylation

Although the protein itself is of molecular mass of 70 kDa, its mature version isolated in vivo has an apparent mass of approximately 130 kDa due to heavy glycosylation. During processing in the ER compartment, several high-mannose oligosaccharide chains are added co-translationally to the core protein. These chains are modified in the Golgi apparatus by mannosidase I and II, and several glycosyltransferases to generate endoH-resistant complex oligosaccharide chains.

The oligosaccharides, while lacking O-glycan modifications, contain sialic acid and oligomers of the N-acetyl-D-glucosamine (Herreman et al., 2003).

Despite of the complexity of glycosylation, the glycosylation is apparently not a requirement for the assembly or activity of γ-secretase. However, the nicastrin ectodomain undergoes a significant conformational change upon incorporation into the γ-secretase complex and the parallel change in the glycosylation can serve as a good indicator of this process (Herreman et al., 2003).

1.1.4.2 The peptidase M20-like domain

The central region of NCT, conferring residues 241-466, bears resemblance to the aminopeptidase M20/M25/M40 superfamily (Fagan et al., 2001). The aminopeptidase family includes, among others, the aminopeptidase from *Streptomyces griseus* and the non-protease transferrin receptor (Figure 1.1.2). Despite a low sequence identity with other members of the aminopeptidase family, in the study of Chavez et al. (2008) algorithms could predict a hydrolase-like fold for the NCT domain with high levels of confidence.
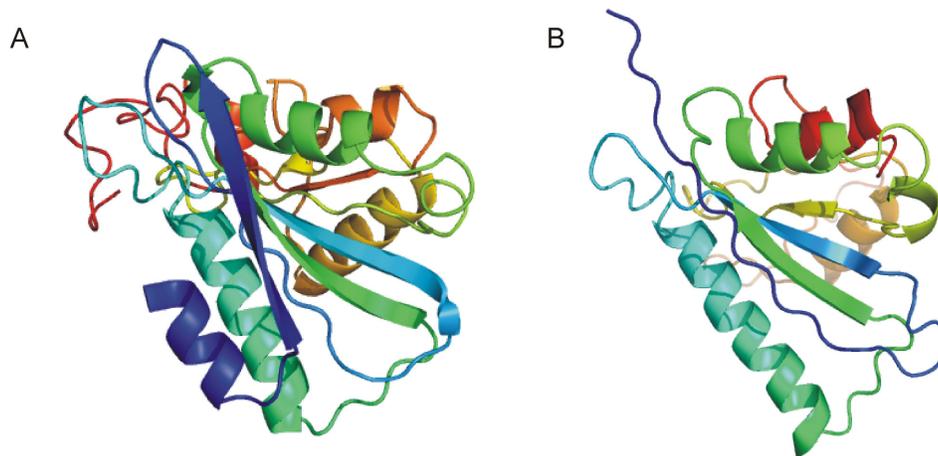
Figure 1.1.2. 3D structures of the domains of aminopeptidase from *Streptomyces griseus* (**A**) and the human transferrin receptor (**B**).

The aminopeptidase-like domain encompasses a conserved stretch of amino acids DYIGS (residues 336–340), which was shown to modulate APP processing. (Yu et al., 2000). Therefore, this domain was dubbed the DAP domain (DYIGS and the peptidase domain).

Localized in the aforementioned domain, the residue Glu333 gained initially much attention, and was suggested to be the major substrate-recognition site (Shah et al., 2005). More recent studies suggest that this residue and the whole DAP domain do not play role in the binding of the N-terminus of the target protein stub. It is, however, involved in the γ-secretase complex assembly on the level of the presenilin incorporation, this function being presumably independent of the aminoacid's charge (Chavez et al., 2008).

1.1.5 Substrates of γ-secretase

1.1.5.1 Alzheimer Precursor Protein (APP)

The β-amyloid peptide (Aβ) is released from its precursor, APP, by two sequential proteolytic cleavages mediated by β- and γ-secretases. β–Secretase (BACE1 - β-site APP-cleaving enzyme 1) removes most of the ectodomain of APP and leaves behind a 99-aminoacid membrane-embedded C-terminus. The APP CTF is subsequently cleaved by γ-secretase to release the Aβ (which is deposited as senile plaques in the brains of patients affected with AD) and the intracellular domain of APP (AICD).

The most popular model suggests that proteolysis of APP occurs in a stepwise manner with γ-secretase cutting APP CTF firstly close to the intracellular border of the membrane (ε-site), generating Aβ48 or Aβ49 peptides (remaining in the membrane) and the AICD, which is released to the cytoplasm. This initiates a

sequential proteolytic cascade, in which further cleavages occur roughly every third amino acid (one α-helix turn) down TMD until the peptide is short enough to be released from the membrane. In this model, Aβ49 is the precursor of Aβ46 (ζ-site), Aβ43, Aβ40 and Aβ37 (γ-sites) while Aβ48 would give rise to Aβ45 (ζ-site), Aβ42 and Aβ39 (γ-sites) (Figure 1.1.3). The two product lines described above do not explain however the generation of Aβ38 (Steiner et al., 2008).
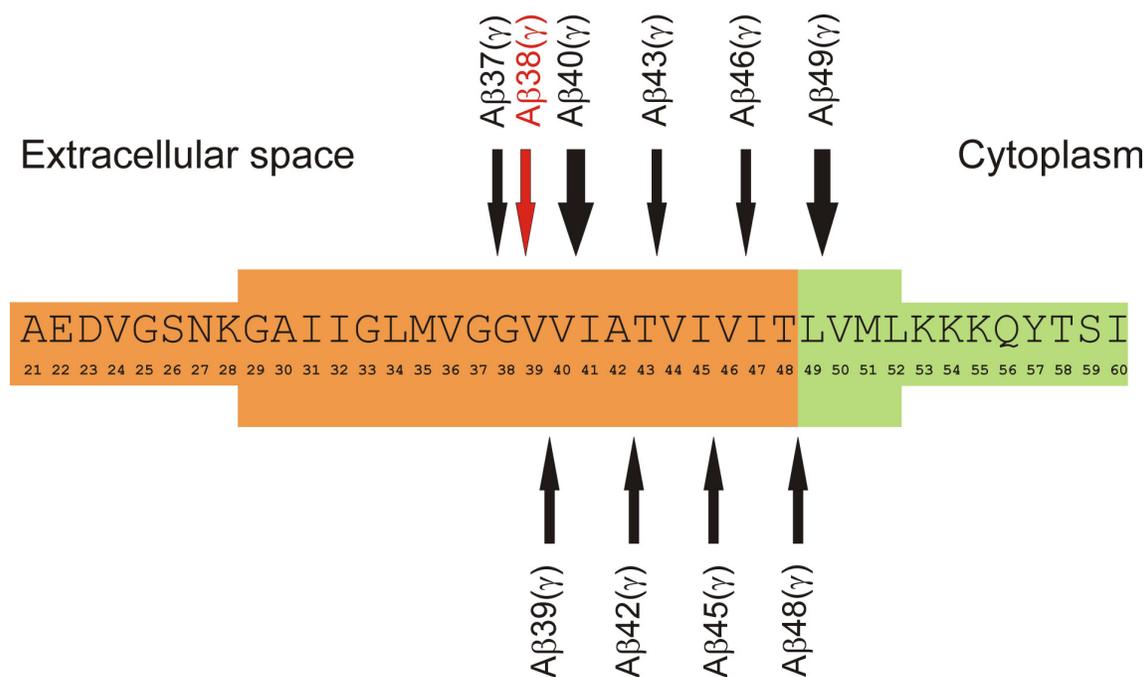
Figure 1.1.3. Scheme of the APP cleavage.

It was also suggested that dimerization of APP is a mechanistic determinant for generation of Aβ42. The dimerization is mediated by the $G_{29}xxxG_{33}$ dimerization motif (one of three existing in APP), which, when disrupted, leads to a selectively reduced production of Aβ42 without significantly affecting the generation of Aβ40 (Munter et al., 2007).

1.1.5.2 Notch

Notch receptors are type I transmembrane proteins that undergo a constitutive furin-like convertase cleavage (S1) in the trans Golgi network. Due to this first cleavage, the mature receptor is presented as a heterodimer on the plasma membrane. The Notch heterodimer consists of the Notch Extracellular Domain (NECD), which is bound in a calcium-dependent manner to a membrane-tethered fragment that includes an extracellular stub, a transmembrane domain, and an intracellular domain. S2 cleavage occurs in response to ligand binding (Delta, Jagged) and releases the majority of the extracellular domain. This cleavage is mediated by proteases of the ADAM/TACE family. The resulting membrane-anchored fragment, referred to as Notch Extracellular Truncation (NEXT), is subject to intramembranous S3 cleavage by γ-secretase, which finally releases NIC. The final cleavage site lies between Gly173 and Val 174 near the cytoplasmic side of the lipid bilayer (Lai, 2002; Kopan and Goate, 2009). Released NICD can then enter the nucleus, where it interacts directly with members of the CSL (CBF1, Su(H), Lag-1) family of transcription factors and participates in transcription activation.

1.1.5.3 Other substrates

Apart from APP and Notch, two best described substrates, γ-secretase cleaves number of other type I membrane proteins engaged in a wide range of cellular processes, including signal transduction (Notch ligands: Delta, and Jagged) cell adhesion (nectin 1a, voltage-gated sodium channel-b2, g-protocadherins, E- and N-cadherins), protein trafficking, tau phosphorylation, and calcium regulation.

1.1.6. Biological significance of γ-secretase

1.1.6.1 Alzheimer disease

Alzheimer's disease, the leading cause of dementia in the elderly, is an irreversible, progressive neurodegenerative disorder clinically characterized by memory loss and cognitive decline, leading invariably to death, usually within 7–10 years after diagnosis. It is characterized pathologically by the presence of senile plaques and neurofibrillary changes in the brains of affected individuals.

The plaques are composed of extracellular aggregates of β-amyloid peptides (Aβ) that are proteolytically produced from Alzheimer precursor protein (APP). Among generated species, Aβ42 is relatively a minor one, however, it is deposited in the Alzheimer disease patients' brains and aggregates much faster than the predominant Aβ40 species.

Several mechanisms have been proposed to explain Aβ neurotoxicity: production of reactive oxygen species such as hydrogen peroxide, nitric oxide, superoxide, highly reactive hydroxyl radicals and nitric oxide; decreased membrane fluidity; energy depletion; alteration of the cytoskeleton; inflammatory processes; and alteration of metal homoestasis. All of these events converge into similar pathways of synaptic disruption, necrosis or apoptosis, leading to progressive loss of specific neuronal cell populations (Masters et al., 2006).

Familial Alzheimer disease has been associated with mutations in APP, presenilin 1 and 2 genes that cause production of increased levels of Aβ42 through the incomplete C-terminal trimming of APP. For example, G384A localized in the GxGD motif on TM7, A434C and P436C mutations localized in TM9, show a significant decrease in Aβ40 with a simultaneous increase in Aβ42 (Tolia et al., 2008).

1.1.6.2 Notch signaling pathway

The Notch signaling pathway of metazoans enables short-range communication between cells. Notch signals are often used to select between preexisting developmental programs. In the context of lateral inhibition, Notch signaling facilitates competitive interactions between cells, where a cell that is adopting a particular fate, signals neighboring cells to inhibit them from adopting the same fate. This mechanism allows adjacent cells with similar potential to adopt different fates, thereby regulating cell diversity in a developing tissue. This occurs in cells throughout development of the organism and during the maintenance of self-renewing adult tissues. Notch signaling is also involved in boundary formation and morphogenesis, and in somitogenesis.

Because Notch plays a critical role in many fundamental processes and in a wide range of tissues, aberrant gain or loss of Notch signaling components has been directly linked to multiple human disorders. These disorders include developmental syndromes (Alagille syndrome, Tetralogy of Fallot, syndactyly, spondylocostal dysostosis, familial aortic valve disease), adult onset diseases (such as CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy), and cancer (Kopan and Ilagan, 2009).

1.1.6.3 γ-secretase as a potential drug target

Specificity of the cleavage performed by γ-secretase is influenced by aminoacid sequence of its subunit, especially presenilin. A presenilin variant bearing mutation F388C on TM6 or I437C in TM9 cleaves Notch, but not APP (Tolia et al., 2006, Tolia et at., 2008). On the other hand, introduction of a phenylalanine at position x of the catalytic GxGD motif in TM7 or substitution of one of TM6 conserved

13

residues (Thr-245, Ser-254, Tyr-256, and Val-261) results in preferential cleavage of APP over Notch (Tolia et al., 2006, Yamasaki et al., 2006).

Due to this variability, γ-secretase seems to be a suitable drug target in the Alzheimer disease therapy, assuming it will be possible to modulate its activity towards APP, without disrupting the Notch signaling pathway.

Classical transition state compounds against aspartic proteases, which target the catalytic site, inhibit similarly all cleavages of all substrates and are therefore unsuitable as potential drugs. However, non-transition state γ-secretase inhibitors, such as peptide-based inhibitors (DAPT), sulfonamides and benzodiazepines, when used at low concentrations, do not affect cleavage at the ζ-site, but are potent inhibitors of Aβ40/42 production (γ-site). Moreover, some non-steroidal anti-inflammatory drugs, such as ibuprofen, sulindac sulfide and flurbiprofen, were shown to specifically abolish the production of the most amyloidogenic Aβ42 species through shifting cleavage specificity towards shorter peptides without affecting NICD production (Tomita, 2007; Tolia and De Strooper, 2009).

1.2 Phycocyanobilin lyases

1.2.1 Phycobilisome

Phycobilisomes are the light-harvesting antennas in cyanobacteria and red algae, which span an absorption range from 460 to 670 nm and transfer excitation energy with high quantum efficiency to the photosynthetic reaction centers. Phycobilisomes are large multiprotein particles of intricate structures composed of intensely colored light-absorbing biliproteins and usually colorless linker proteins. Two structural and functional regions can be distinguished within these large complexes namely a thylakoid membrane-attached core and a series of rodes protruding from it (Figure 1.2.1).
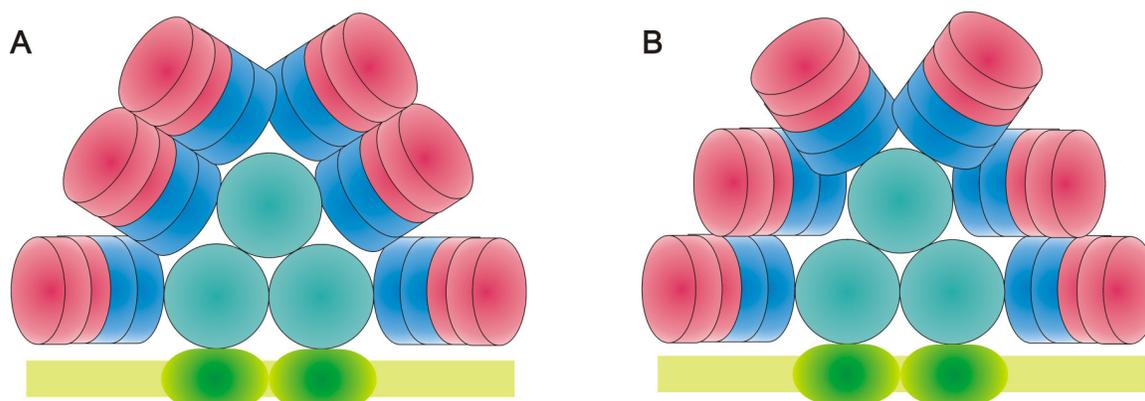
Figure 1.2.1. Models of phycobilisome according to: (**A**) MacColl (1998) and (**B**) Jiang et al., (2001). PE is colored red, CPC - blue, APC - green-blue, photosystem II - green, thylakoid membrane - yellow.

1.2.1.1 Phycobiliproteins

Biliproteins can be divided into four groups: phycoerythrins (PE), phycoerythrocyanins (PEC), phycocyanins (CPC), and allophycocyanins (APC). All biliproteins function as heterodimers built from α- and β-subunits; this complex

through convention, and hereafter, is being referred to as the "monomer". Each subunit contains eight helices, with six of them (helices A-F) folding into a globin-like structure, while additional two (helices X and Y) participate in the monomer formation (Figure 1.2.4A). The monomer is assembled through interaction of helices XY of one of subunits with the globin-like domain of the other (Anderson and Toole, 1998).

Each subunit contains at least one open-chain tetrapyrrol prosthetic group, called phycobilin, to which it owes its spectroscopic properties. Commonly incorporated bilins: phycoerythrobilin (PEB), phycocyanobilin (PCB), phycoviolobilin (PVB), and phycourobilin (PUB) are shown in Figure 1.2.2.
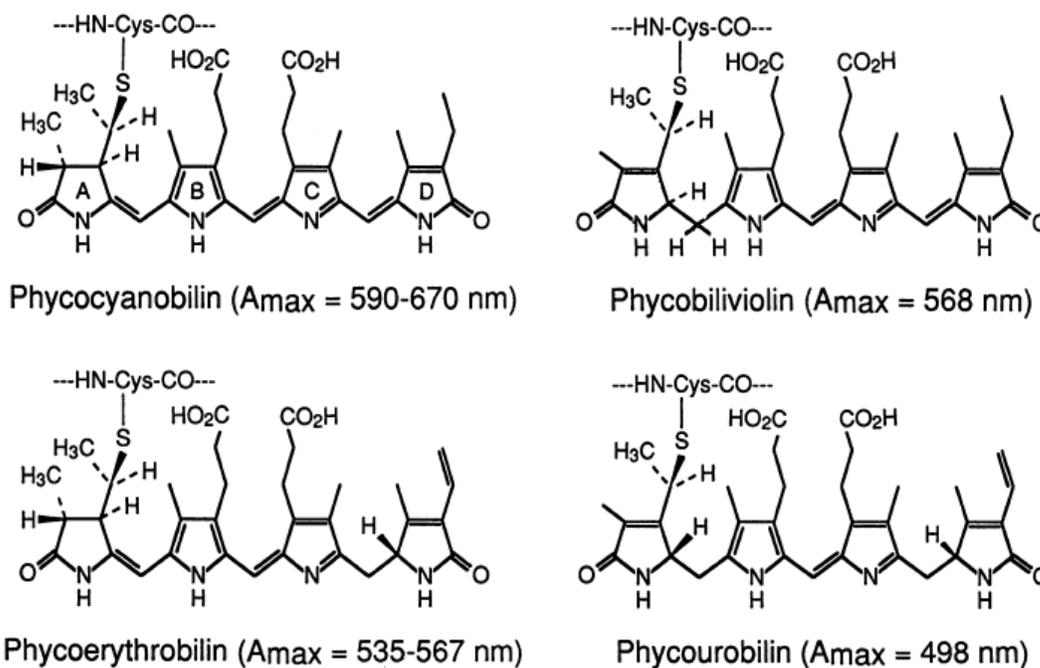
Figure 1.2.2. Structures of phycobilins. Absorption maxima are given for the chromophore in a form bound to the protein.

Bilins are covalently attached to a cysteinyl residue of an apoprotein by a tioether bond via ring A (and sometimes D). Two bilins are invariably attached to Cys84 (using consensus numbering) of each subunit and additional chromophores

may be linked to other cysteines found on subunit ß (most commonly Cys155 in CPC, PEC, and PE). When bound to proteins, bilins remain in an extended conformation, what, together with the influence of environment, changes their spectroscopic properties. Schematic representations of phycobiliproteins together with their absorption maxima are shown in Figure 1.2.3.

Phycoerythrin λ_{max}= 560 nm   Phycoerythrocyanin λ_{max}= 575nm

α   β   α   β

PEB   PEB PEB   PVB   PEB

PEB   PEB (PUB)   PEB

Phycocyanin λ_{max}= 620 nm   Allophycocyanin λ_{max}= 652 nm

α   β   α   β

PCB   PCB   PCB   PCB

PCB
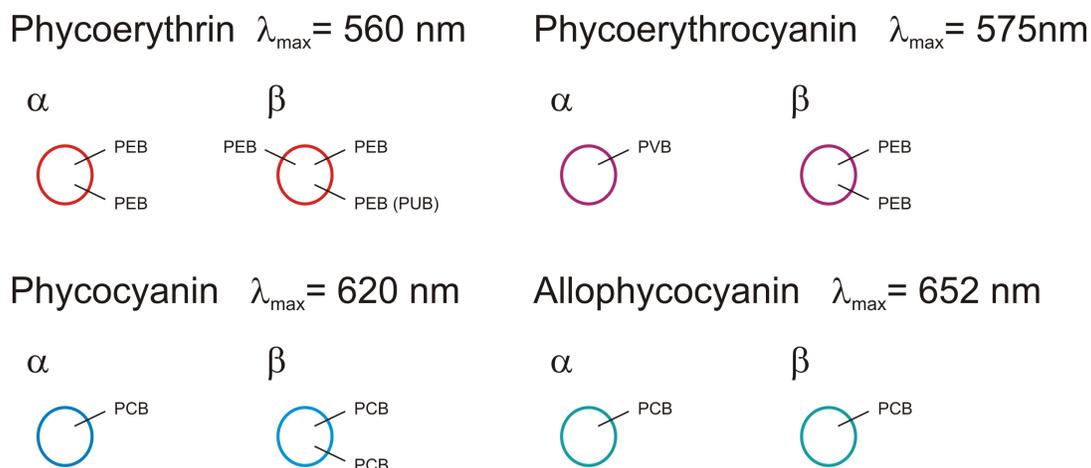
Figure 1.2.3. Schematic representation of phycobiliproteins. α- and β-subunits of apoproteins are shown as rings with coloration corresponding approximately to the true proteins' color.

Phycobiliproteins are assembled into the discs composed of three monomers. The proteins are arranged with the α-subunit of one monomer being in proximity of the β-subunit of the neighbouring one. Such a trimer is 110 Å wide and 30 Å thick with a central channel of 35 Å in diameter (Figure 1.2.4). Trimers associate in the face-to face fashion forming hexamers (double discs).

A detailed phycobiliprotein structure will be described for PEC from *Mastigocladus laminosus*, as this protein was used in parts of this study. In the β-subunit, one chromophore is embedded between helices E, F' and G, while the second one lies more peripherally, contacting helix A and the loop between helices G and H (Figures 1.2.4 and 5). The α-subunit of PEC shows high similarity to its β-counterpart. The main differences in the secondary structure are related to the
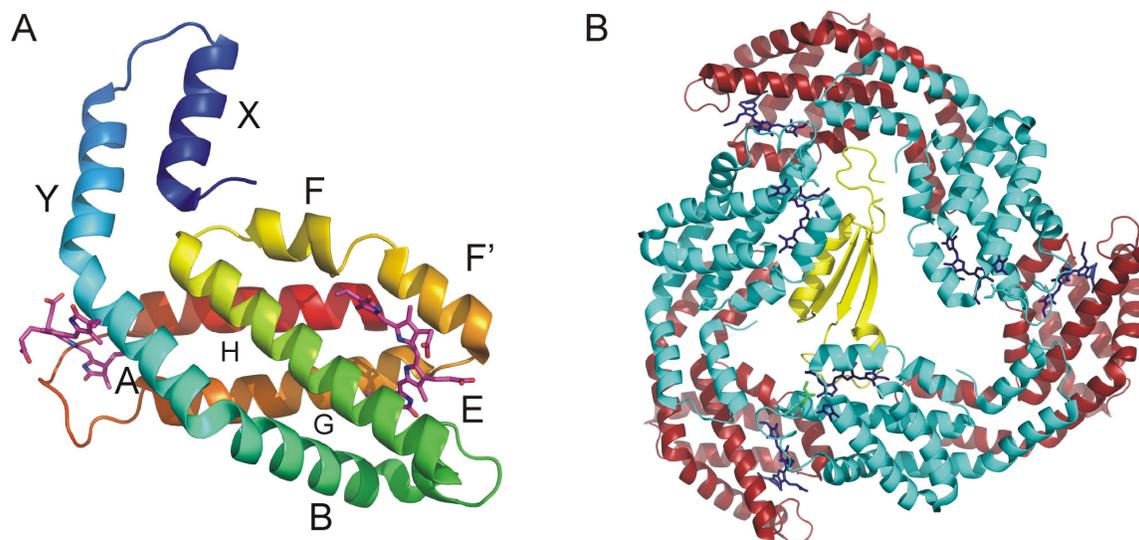
Figure 1.2.4. (**A**) Globin-like fold of an exemplary cyanonbilin (β-subunit allophycocyanin) (**B**) Trimeric disc of allophycocyanin with a linker. Subunits α- are colored red, β- are colored cyan, the linker protein is colored yellow. Chromophores are presented in heavy atoms representations.

chromophore binding: the bilin-interacting loop between helices F' and G is significantly shortened in PecA, while the fragment between helices E and B is more pronounced (Schimdt et al., 2006). Flexibility of this longer loop is an important factor in the unique photochemistry of α-PEC. Even though all phycocyanobilins are capable of adopting two conformations: E and Z forms through twisting around the Δ15,16 double bond between the C and D pyrrole rings (Figure 1.2.2), the E form was detected exclusively in PVB. It is proposed that due to its flexibility, the loop connecting B and E helices can stabilize the E form of PVB (Wiegand et al., 2002).

It was previously observed in our laboratory that prolonged incubation of α-PEC yields a degradation product stripped off the first 36 aminoacids. This fragment, named mini α-PEC, is stable and displays full photochemistry of α-PEC and was therefore subjected to the structural and functional analysis (unpublished results).
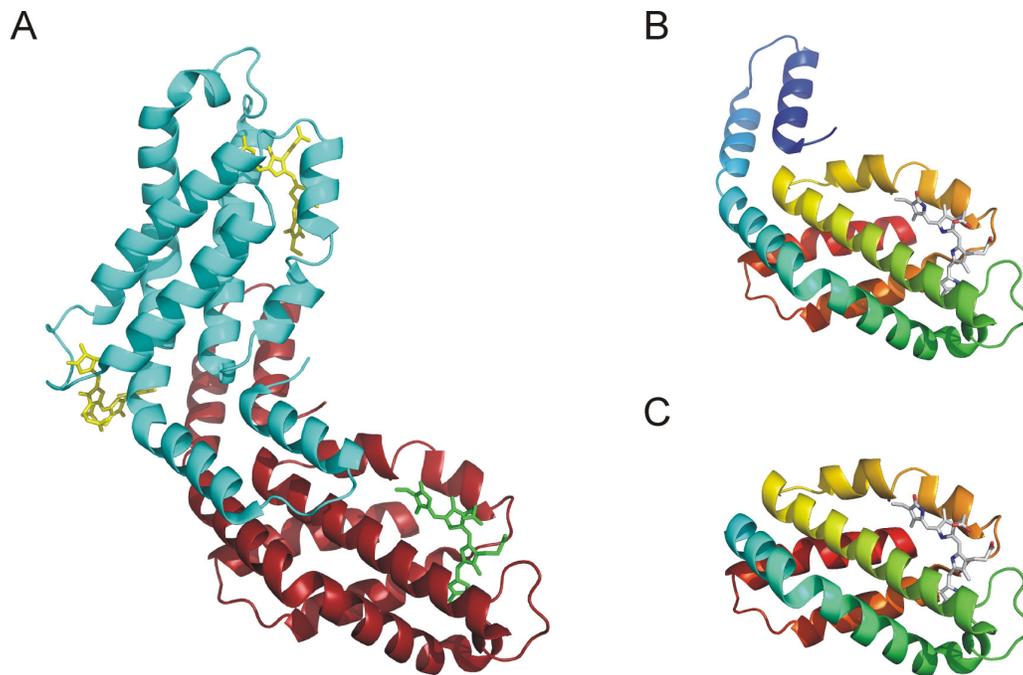
Figure 1.2.5. Structure of PEC from *Mastigocladus laminosus*. (**A**) Heterodimer (subunit α colored red, subunit β colored cyan, PCB colored yellow, PVB colored green). (**B**) α-PEC. (**C**) Model of mini α-PEC.

1.2.1.2 Organization of phycobilisome

Both core and rodes are composed of so-called cylinders; each cylinder being a stack of hexameric double discs arranged one after another.

The structure of the core varies between organisms with the number of cylinders ranging from two to five. When only two cylinders are present, they are arranged in one layer, both in contact with the tylakoid membrane. Additional cylinders are packed hexagonally in a second (or third) layer. Each cylinder consists of four trimers of allophycocyanin (APC) that gives it the length of 120 Å.

Invariant in diameter (about 110 Å) cylinders of rods are longer than these of the core, consisting of three hexamers. Protein composition of rods varies dependent on the distance from the core. Phycoerythrin (PE), accompanied in some species by

phycoerythrocyanin (PEC), can be found at the periphery of the structure, while phycocyanin (CPC) in the proximity to the core linking PE with APC.

It is accepted that six to eight rods form a fan-like assembly radiating from the core, however, their exact spatial arrangement remains a matter of discussion. Several models have been proposed: one describes rods arranged in three pairs surrounding the core at right angles (Figure 1.2.1A), while the other suggests two single rods forming angle of about 45° to the remaining rods assembled in pairs (Figure 1.2.1B).

Additionally to the pigmented phycobiliproteins, the phycobilisomes contain non-pigmented polypeptides that have been named linker proteins. They participate in assembly processes, stabilize interactions between hexamers, and modulate the absorption characteristics of the phycobiliprotein.

1.2.1.3 Energy transfer within the phycobilisome

Phycobilisome serves as the antennae for the photosystem II expanding the range of light energy that can service photosynthesis to 450-650 nm. The efficiency of energy transfer reaches 90% in intact cells.

The mechanism of energy migration through the phycobilisome is called resonance energy transfer. It occurs by a donor molecule being excited, going to the lowest vibrational energy level by rapid vibrational relaxation, and then a radiationless energy transfer to the acceptor molecule. The resonance energy transfer requires the spectral overlap of the donor fluorescence and acceptor absorbance spectra and a specific distance and relative geometry of the donor / acceptor chromophore pair. All of this is accomplished by arrangement of chromoproteins in phycobilisome, with PE absorbing the green light on the periphery and CPC absorbing the orange, closer to the core. The bilin network corresponds to a funnel with diminishing number of

chromophores closer to the core (30 bilins for a PE hexamer, 18 bilins for a CPC hexamer, 12 bilins for APC hexamers). The physical arrangement of the different phycobiliproteins creates a clear energy gradient from PE (or PEC) through PC to APC, and down into the chlorophyll of the reaction center.

After being absorbed by protein in the rod, energy quantum migrates to APC of the core and afterwards reaches the terminal energy acceptor - linker protein L_{CM}. L_{CM} is the membrane-embedded protein which contains a single chromophore that is red-shifted compared to the other biliprotein (absorption maximum 680 nm). This enables the energy transfer directly to chlorophyll. L_{CM} can be found only in those core cylinders that are in contact with the thylakoid membrane

1.2.2. Biosynthesis of biliproteins

1.2.2.1 Biosynthesis of bilins

Cyanobilins are derivatives of protoheme. Through the action of the heme oxygenase, the macrocycle of protoheme is cleaved resulting in an open chain tetrapyrrol. This compound is reduced by the ferredoxin-dependent bilin reductase in three possible positions (the end product depending on the reduction site). Two types of free cyanobilins are found in cyanobacteria: PCB and PEB. Others, PVB and PUB, although present in forms bound to protein, are created simultaneously to the attachment process (Scheer and Zhao, 2008).

2.2 Lyases

Although most of cyanobiliproteins are able to spontaneously attach a chromophore, this process is characterized by low fidelity and results in a mixture of products showing oxidation and incorrect stereochemistry. The only known exception is the L_{CM} protein, which is able to autocatalically and correctly attach PCB yielding

the active end product. Synthesis of all other cyanobilin relies therefore on the action of phycocyanobilin lyases. Several groups of these enzymes have been characterized so far, differing in the substrate specificity and degree of homology.

When considering the attachment specificity, four different groups of cyanobiliprotiens (PE, PEC, CPC, and APC, each containing α- and β-subunits), four bilins (PEB, PCB, PUB, and PVB) and four attachment sites (Cys84, Cys155 present in all cyanobiliproteins, plus additional Cys48, Cys59 in CpeB) must be taken into account.

1.2.2.2.1 S- and T- type lyases

Enzymes belonging to this family are responsible for attachment of the chromophore to majority of biliproteins (Table 1.1). The type S lyases have broad range of substrates for both the apoprotein and the bilin. Their action is restricted, however to the Cys84 position (irrespectively of the subunit).

The lyase CpeS from *Anabaena* PCC7120 and the CpcS *Synechococcus* PCC7002 show no discrimination to protein substrate, acting on members of every group of cyanobiliproteins (Table 1.1.1). The latter requires for its activity a related protein, CpcU, forming a heterodimeric complex.

The Cys155 site in CpcB and PecB in *Synechococcus* PCC7002 acquires its chromophore thanks to the activity of the CpcT lyase. However, *in vitro*, PCB can be attached correctly to both cysteines only if CpcT is followed by CpcS because presence of PCB at Cys84 interferes with the attachment at Cys155 (Zhao et al., 2007b).

It is still unknown which lyases are responsible for the attachment of the chromophores to sites Cys48 and Cys59 in PE.

Table 1.1.1

Lyase	Species	Chromophore	Apoprotein	Binding Site	Reference
CpcS/U	*Synechococcus* PCC7002	PCB	ApcA, ApcB	84	Saunee et al., 2008
			PecB CpcB		Shen et al., 2006, 2008
		PEB	CpeA CpeB		
CpeS	*Anabaena* PCC7120	PCB	CpcB PecB	84	Zhao et al., 2006a
		PCB	ApcA ApcB	84,155	Zhao et al., 2007a
		PEB	CpeA CpeB	84	Zhao et al., 2007a
CpcT		PCB	CpcB PecB	155	Zhao et al., 2006a
PecE/F	*Mastigocladus laminosus* PCC7603, *Synechococcus* PCC7002	PCB → PVB	PecA	84	Zhao et al., 2002
Cpc E/F	ubiquitous	PCB	CpcA	84. And reverse reaction	Zhou et al., 1992
		PCB	CpeA	84	Kahn et al. 1997

1.2.2.2.2 E/F-type lyases

Chromophylation of the sole binding site on the α-subunit (Cys84) on CpcA and PecA is catalyzed by the E/F-type lyases: CpcE/F and PecE/F, respectively. Lyases from this family work in heterodimeric complexes and are usually found downstream to their substrate protein in the same operon. CpcE/F-type lyases, abundant among cyanobacteria, catalyze attachment of PCB to CPCA and CpeA.

The PecE/F lyases catalyze a unique reaction of the chromophore attachment combined with isomerization of the Δ4,5 double bond of the PCB chromophore to the Δ2,3 position, thus creating phycoviolobilin (PVB). This reaction is rare among the

cyanobacteria and only few isomerizing lyases have been identified. PecE/F from *Anabaena* PCC710 and, the best studied, PecE/F from *Mastigocladus laminosus* PCC7603, attach PCB to PecA. Another member of this group, a protein from *Synechococcus* WH8102 seems to be a fusion of two lyase's subunits (E and F). It is hypothesized that the substrate of this lyase is PEB, which is isomerized to PUB during the attachment to the apoprotein (Six et al., 2005).

PecE/F subunits form a relatively stable complex and both partners are capable of binding, although weakly, the bilin substrate. Since the PecE subunit bears more resemblance to respective CpcE proteins, it is suggested that it is the subunit F which possesses the isomerizing activity (Zhao et al. 2005). PecE, on the other hand, if present alone, catalyzes stereochemically correct attachment of PCB to PecA, but the chromophore remains unisomerized in this process. Product of this reaction is a stable adduct, called P641, having a high, yet different from the natural PEC, fluorescence and absorbance. Activity of PecE in this process was classified as a chaperone-like assistance in the autocatalytical attachment of PCB through control of conformation of the free bilin. At the same time, the ternary complex PCB:PecA:PecE can be considered as a stable intermediate of the natural attachment reaction (Boehm et al., 2007).

Deletion studies of *Mastigocladus laminosus* PecE/F gives some insight into its residues´ function. The C-terminus of PecF seems not to be crucial for its activity, since mutants covering residues 1-201 and 1-191 retain moderate catalytic activity. In the case of PecE, the complete C-terminus (except for the last three aminoacids) is essential for the enzymatic activity (Zhao et al., 2005).

The recently proposed mechanism of the E/F-type lyases features a nucleophilic attack at the central methine bridge (Figure 1.2.6B). The nucleophile

addition produces mixture of isomers with differently placed double bonds bound to

the enzyme via the cysteine or histidine residue (Figures 1.2.6C and E). At this step,

the pathways of reactions catalyzed by isomerizing and non-isomerizing lyases, part.

CpcE/F are selective for adducts shown in the Figure 1.2.6C from which the

chromophore is transferred to the acceptor protein with the concomitant back

isomerization. Isomerizing lyases, on the other hand, favor formation of products

shown in Figure 1.2.6F which are transferred to the apoprotein without further

changes. According to this model, the chromophore is isomerized by all of the lyases,

transiently with the "nonisomerizing" lyases (CpcE/F), and irreversibly with the

isomerizing ones (PecE/F) (Tu et al., 2009).
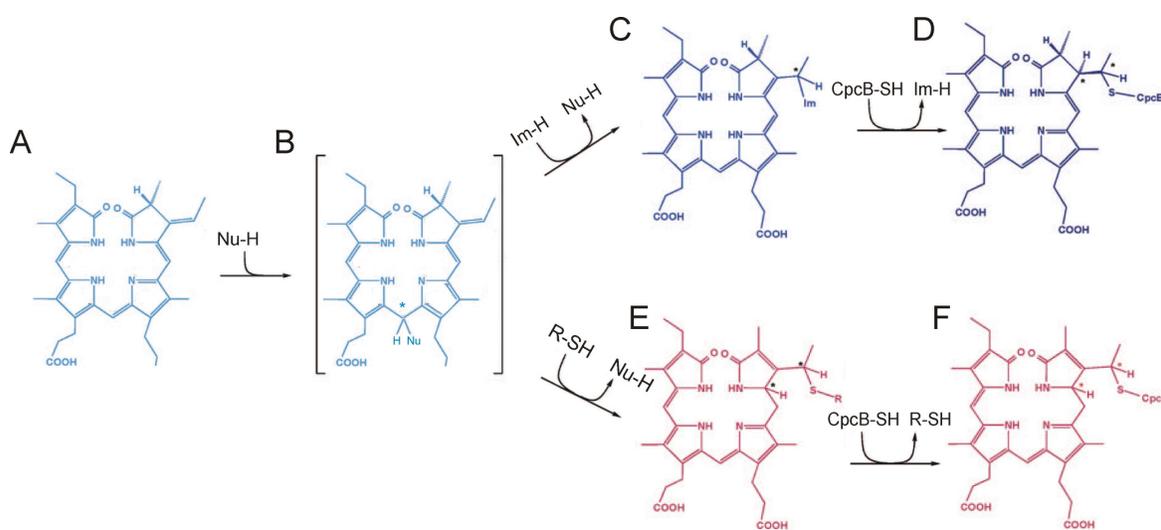
Figure 1.2.6. Products of the nucleophilic addition to PCB (Tu et al., 2009).

1.3. Site-specific integrases

1.3.1 Site-specific recombination

Site-specific recombination is a process of DNA breakage and re-ligation that requires neither DNA synthesis nor high-energy cofactors and produces newly arranged DNA segments. Site-specific recombination is used as a mean of physical joining or separating DNA fragments by organisms, as well as genetic elements. It is utilized in phage integration and excision, cointegrate resolution, DNA transposition, gene expression regulation, and last but not least, as a mean of generating genetic diversity through acquisition of advantageous genes or gene segments (Sadowski, 1986).

Depending on the initial arrangement of the parental recombination sites, site-specific recombination has three possible outcomes: integration, excision, or inversion. Integration results from recombination between sites on separate DNA molecules (provided that at least one of the parental chromosomes is circular). For sites located on the same chromosome, the outcome is determined by their relative orientation. If recombination sites are in a head-to-tail orientation the process will end with excision, while head-to head orientation will result in inversion (Figure 1.3.1).

Site-specific recombination involves two DNA fragments sharing a stretch of identical bases and a specialized recombinase protein (responsible for recognition, breaking and rejoining of DNA). These elements, along with the detailed mechanism of reaction, will be described below.
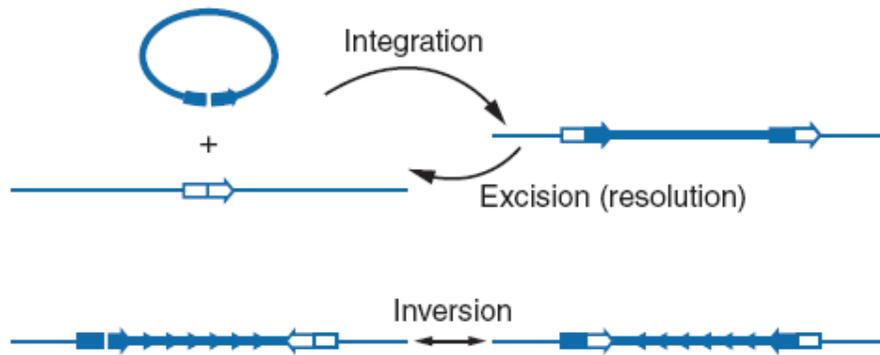
Figure 1.3.1. Three possible outcomes of site-specific recombination: integration, excision and inversion.

1.3.1.1 Structure of the binding site

Site specific recombination requires two DNA fragments, called attachments sites (att), which consists of centrally placed, short segments of residues identical at both sites (the so called overlap region, O), flanked by inverted repeats (called core-binding sites). In the case of phage integration, core-binding sites are labeled C and C' when building a phage attachment site (attP) or B and B' when placed on the bacterium chromosome (attB). Recombination process produces attachment sites with rearranged elements: the left attachment (attL) can be presented as BOC', while the right one (attR) as COB' (Figure1.3.2A).

This simple structure of the attachment site is present in systems like Cre from phage P1 (Van Duyne, 2001) or FLP from the 2μ plasmid of *Sacharomyces cerevisae* (Shaikh and Sadowski, 1997). Many phages, however, utilize more complicated mechanisms. Apart from the core-type sites, they posses multiple adjacent, so-called arm-type sites, capable to interact with integrases, cofactors, and host-derived factors (Figure 1.3.2B). The purpose of binding of these proteins may be regulatory, structural, or both. They may initiate, or stabilize the pairing of

recombination sites, or inhibit inappropriate pairings; they may deliver the recombinase catalytic domains to the crossover site, activate the recombinase or determine the directionality of recombination (Ross and Landy, 1982).
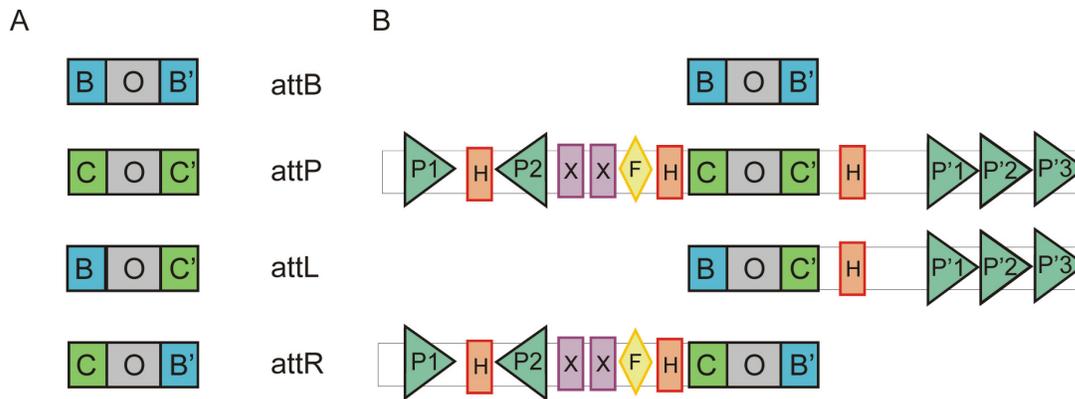
Figure 1.3.2. Comparison of simple (**A**) and complex (**B**) attachment sites. O-overlap region; C – phage core-type binding sites; B – bacteria core-type binding sites; P - arm-type binding sites; H, F, X – binding sites for cofactors: IHF, Xis, and Fis, respectively.

The P arms of the phage lambda contain several arm-type binding sites (two on arm P and three on arm P'). Different pairs of attachment sites are used for integration (attP and attB) or excision (attL and attR) with resultant differences in the relative orientations of P and P' arm. P1, P'2 and P'3 sites are used during integration, while remaining ones during excision.

1.3.1.2 Integrases

Site-specific recombinases can be divided into two families, the tyrosine recombinases and the serine recombinases, named after the amino acid residue that forms a covalent protein-DNA linkage in the reaction intermediate. Their mechanisms of action are distinctly different, with tyrosine integrases breaking and rejoinining single strands in pairs to form a Holliday junction while serine recombinases cutting

all strands in advance and performing exchange and religation afterwards. Only the tyrosine family will be described here in detail (Grindley et al., 2004).

Simpler integrases, like Cre and Flp (from systems lacking arm-type sites), consist of two, mostly helical domains: a core-binding domain and a catalytical domain connected by a short, helical coupler. Integrases originating from more complex systems (like the phage lambda integrase) contain an additional, N-terminal domain responsible for binding of arm-type sites.

Structure of the active site of tyrosine integrases is conserved among members of the family. It contains four, almost absolutely conserved residues: two arginines and one lysine along with a tyrosine nucleophile. Together with two additional (less conserved) residues, namely lysine and histidine (in some cases substituted by tryptophane), they make a RKHRH(W) pentad. These aminoacids cluster in space around the catalytic tyrosine composing the active site (Figure 1.3.3) (Groth and Calos, 2004).

During the cleavage, the hydroxyl of tyrosine nucleophilically attacks the phosphate backbone of DNA. The 3' end of the broken DNA strand is covalently attached to the catalytical tyrosine while the 5' end remains free to subsequently attack the protein-DNA bond of the site of synapsis. Ligation of strands from opposite att sites results in the Holliday junction.

Conserved arginines coordinate the non-bridging oxygens of the scissible phosphodiester stabilizing the pentacoordinate transition state. Histidines (or histidine and tryptophane) may act as general acid and general base. Their character interchange upon cleavage – ligation reactions step. During cleavage, first histidine acts as general base through removing proton from tyrosine while second histidine protonates the liberated 5' end of the product DNA acting as a general acid. These

roles reverse during the ligation (Grainge 1999). According to other models, the general acid role is fulfilled by lysine residue (Grindley et al., 2006).
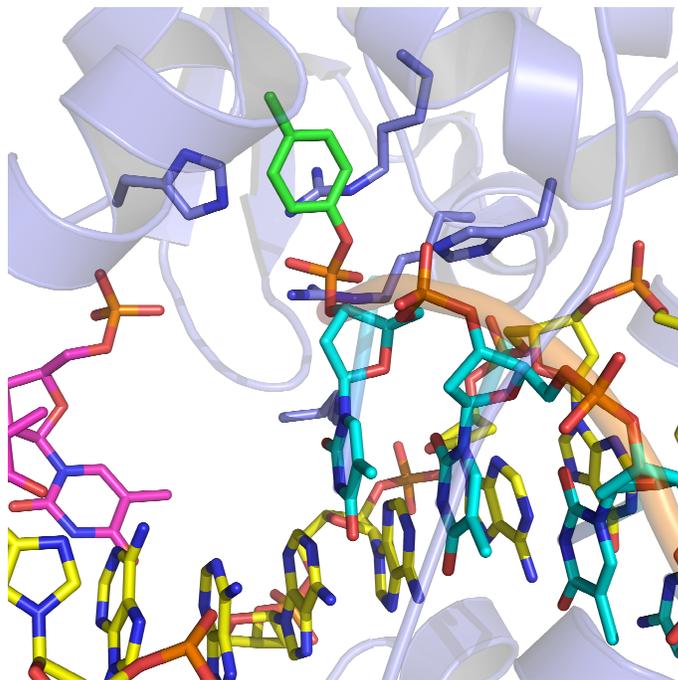
Figure 1.3.3. Active site of the lambda integrase. Conserved residues are shown in stick representations. Cleaved DNA strand is colored cyan (the protein-bound part and magenta – free part). Uncleaved DNA is colored yellow.

Unlike catalytic and core-binding domains, the arm-type binding domains are structurally less conserved. The best-studied N-terminal domain, that of phage lambda, is shown in Figure 1.3.3. Its major groove interacting surface is composed of a three-stranded β-sheet. Upon DNA-binding, the N-terminus undergoes structural organization and forms a short α-helix which projects into the minor groove (Fadeev et at., 2009). In the transposone 916 integrase, the three-stranded β-sheet motif is conserved, however DNA interactions are dependant also on the long loop connecting the sheet and C-terminal α-helix (Connolly et al., 1999; Wojciak et al., 1999b).

The integrase binds weakly to the core DNA and, consequently, relies on binding to the high-affinity arm DNA sites for its delivery to core sites (Radman-Livaja et al., 2005).

A   B
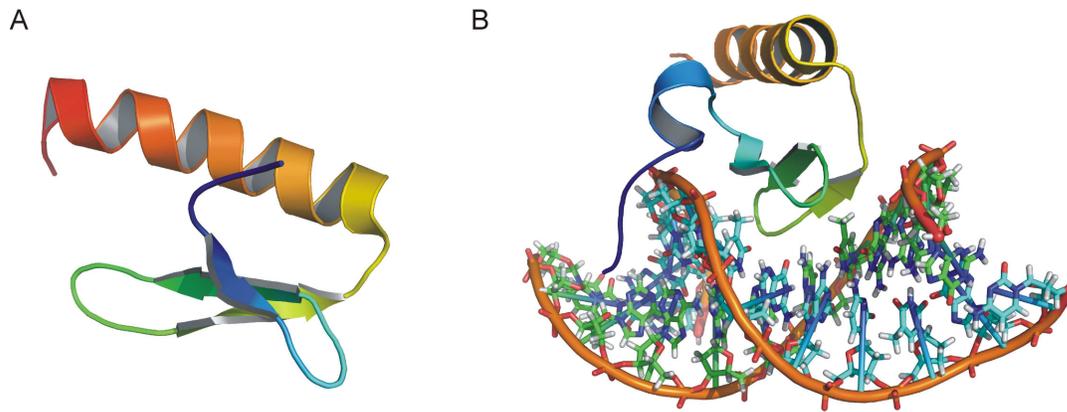
Figure 1.3.4. The arm-type binding domain of phage lambda without (**A**) and with the bound arm-type attachment site (**B**)

1.3.1.3 Structure of the Holiday junction

Each core-type binding site binds one monomer of the integrase, inserting between its core-binding and catalitical domains. When this arrangement is repeated on all four core-binding sites taking part in recombination, a tetrameric structure is formed. The tetramers have twofold and pseudo fourfold symmetry, and hold the Holliday junction intermediate in a nearly square planar conformation (Figure 1.3.5). Catalytic domains of integrases interact with one another in a skewed fashion by swapping parts of their C-terminus with neighboring monomers. This exchanged fragment follows the reactive tyrosine-bearing helix. In Flp recombinase the tyrosine's helix itself is swapped; this supplies each monomer with an active site from a neighboring partner (such arrangement of the active site is called *trans*) (Chen et al., 2000; Martin et al., 2002).

In a tetrameric synaptic complex only two, diagonal monomers are active at the same time (what leads to the true twofold symmetry and only approximate

fourfold one). The alternating interfaces between subunits directly affect the active site conformations, with only one type of special arrangement of the swapped C-teminus allowing proper positioning of catalytic tyrosine.

The arm-binding domains lie in an additional, intertwined layer of the protein complex displaying the twofold but not pseudo fourfold symmetry. Each N-terminal domain is positioned below a core-binding domain belonging to a neighboring monomer and not that of its own. The arm-type binding domains from the monomer lying on the opposite site of the synaptic complex are bound to one arm of DNA (Biswas et al., 2005).
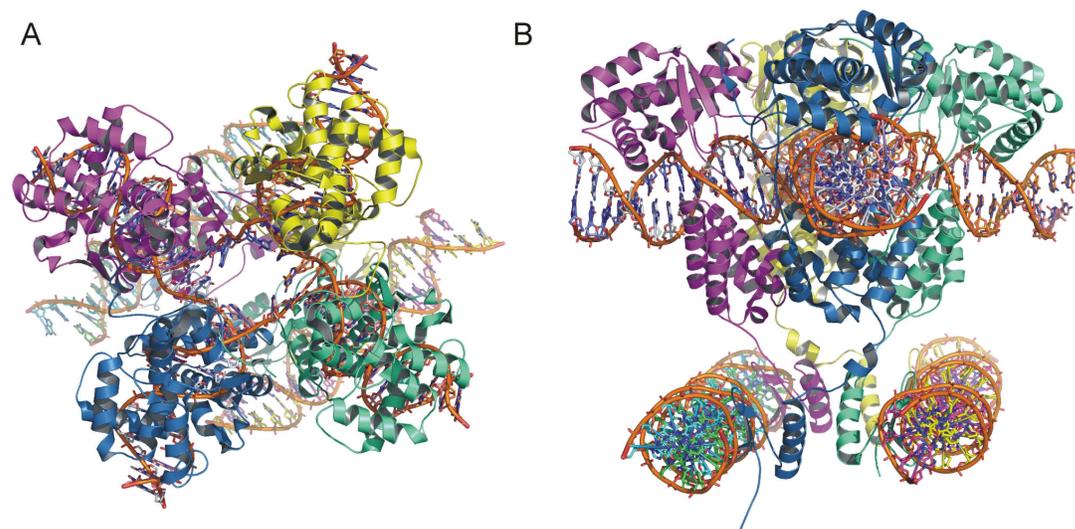
Figure 1.3.5. Lambda integrase bound to the Holliday junction, top view (**A**) and side view (**B**).

1.3.1.4 Mechanism of action

Recombination is initiated when one strand of each duplex is cleaved by a nucleophilic tyrosine, creating the covalent DNA-protein link between tyrosine and the 3' ends of the DNA. Freed 5' end of the DNA strand migrates towards active site of the integrase monomer bound to the opposite side of synapsis. The next step involves exchange of three bases and attack of the free hydroxyl at the 5' end on the

phosphotyrosine link DNA-protein linkage and joining two DNA strands from the opposite attachment site, which forms the Holiday junction. The complex can then isomerize so that the inactive monomers become active and vice versa. This enables a repeat of the whole process; i.e., the second, untouched strand is attacked, and the new 5' end migrate over and attack their partners' 3' phosphotyrosine linkages, freeing the protein, resolving the Holliday junction, and completing the reaction (Figure 1.3.6).

In order to establish correct interactions between monomers, the DNA is bent within the spacer region. The direction of the bend affects proteins' interfaces and determines which strand will be cleaved as first. Binding of the arm sites constrains the topology of the bound DNA and promotes isomerization towards a conformation that is proficient for the cleavage of the second strands to and completion of recombination.
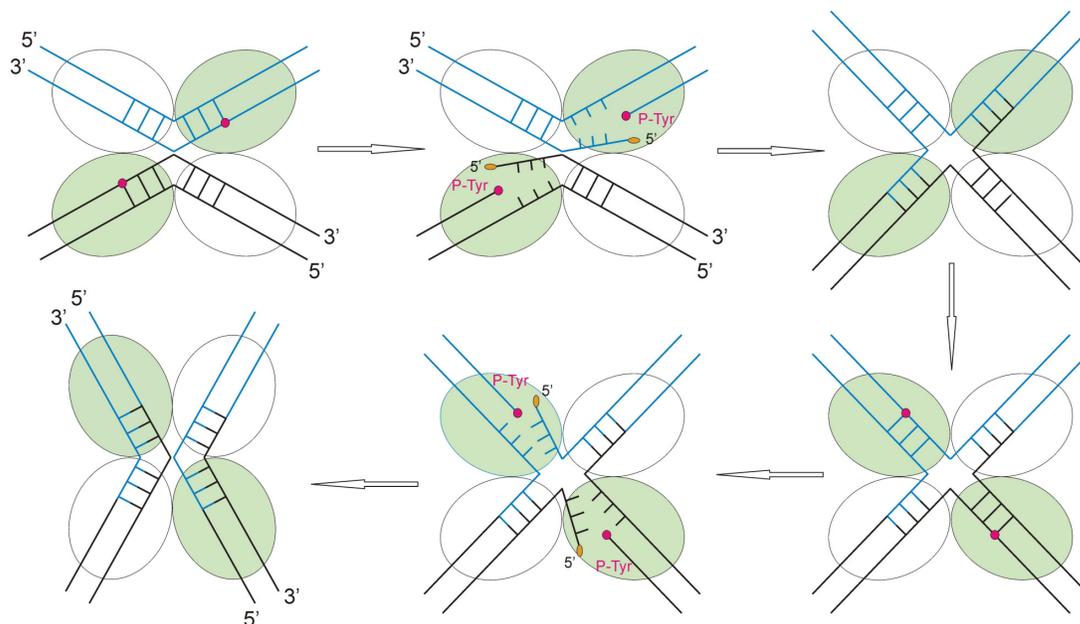
Figure 1.3.6. Mechanism of site-specific recombination. Active subunits are colored green.

1.3.2. Horizontal gene transfer

Apart from the core genes encoding essential cellular functions, bacterial genomes also harbor regions of large foreign DNA coding accessory genes that supply bacterium with adaptative traits improving survival and rapid spread within a niche. These genes form a "flexible gene pool" and are thought to be acquired by the horizontal gene transfer (HGT). Horizontal gene transfer is defined as the transfer of genetic material between bacterial cells uncoupled with cell division and it enables the exchange of genetic material between different prokaryotic species or even genera. In prokaryotes, three principal mechanisms facilitate HGT: natural transformation, conjugation, and transduction. Transformation is the process of acquiring free DNA from bacterium's surroundings. Conjugation involves transfer of conjugative plasmids through a designated tube-like structure - a pilus, from a donor to a recipient bacterium cell. Transduction is the movement of genes from one prokaryotic organism to another via viruses (Gal-Mor and Finlay, 2006).

Laterally transferred DNA comprises phages, plasmids, "simple" and conjugative transposons, integrons, smaller genomic islets (< 10 kb) and genomic islands (GEIs). Genomic islands are defined as relatively large (10 - 200 kb) segments of DNA, usually bearing the GC content and codon usage different from the rest of bacterium genome, that carry genes offering a selective advantage for the host bacteria. They are often flanked by 16-20 nucleotide-long direct repeats (DRs), which arise by site-specific integration of the island into the host genome (integration site being often a tRNA gene). Since most GEIs are, or used to be, mobile, they code the sequences that participate in the transfer, (like the attachment sites and integrases genes or factors related to plasmid conjugation systems). The genomic

islands' coding capacity is very diverse, including such traits as symbiosis, new metabolic pathways, antibiotic resistance, and virulence (Juhas et al., 2009).

1.3.2.1 Yersinia High Pathogenicity Island

Pathogenicity islands represent a distinctive subset of GEIs comprising specific gene clusters, which code for bacterial virulence factors, e.g. adhesins, iron-scavenging siderophores, capsules, endo- and exotoxins, and type III and type IV secretion systems. Such islands have been found in a wide variety of bacterial species including *Escherichia coli, Salmonella, Helicobacter pylori, Dichelobacter nodosus*, and *Vibrio cholerae* (Bedenek and Schubert, 2007).

One of the most widely distributed PAIs is the *Yersinia* High-Pathogenicity Island (HPI), first described in the plague agent *Yersinia pestis* and afterwards detected in other *Yersinia* species: the *Y. enterocolitica* biotype 1B and *Y. pseudotuberculosis*, as well as in other members of the family *Enterobacteriaceae*, including *Escherichia*, *Klebsiella*, *Enterobacter*, *Citrobacter*, *Salmonella*, and *Serratia.* It displays a highly conserved genetic organization in all these genera, which might indicate the recent acquisition and distribution of the island, and thus provides a good model for studying PAI evolution and transfer mechanisms. HPI's presence is essential for the expression of a high-virulence phenotype.

Y. enterocolitica invades gut epithelial cells and causes severe, systemic disease on entering the lymph system. *Y. pseudotuberculosis and Y. enterocolitica* are enteropathogenic bacteria transmitted orally and responsible for intestinal symptoms characterized by a mesenteric lymphadenitis and a terminal ileitis, respectively. *Y. pestis* is the cause of the bubonic and pneumonic forms of plague, one of the most feared and deadly infectious diseases, which, having killed millions of people in worldwide epidemics, is now classified as a re-emerging disease with 2000

cases reported per year (Carniel, 2001; Titball et al., 2003). *Y. pestis* bacteria are regurgitated into the human blood stream following bites from infected fleas (carrying the pathogen from rodents, the primary hosts). Since almost identical genetically HPI is present in *Y. pseudotuberculosis*, as well as in *Y. pestis*, pathogens causing drastically different clinical and epidemiological features, it is clear that the HPI is necessary but not sufficient to cause severe infections in humans. Additional bacterial factors (like the pYV virulence plasmid of *Y. pestis*) are required to establish a life-threatening disease (Bach et al., 1999).

Functional trait carried by HPI is an iron scavenging system mediated by yersiniabactin, a chelator molecule, which has the ability to entrap iron from the environment. Acquisition of iron is an essential requirement for the microorganisms in the host, however, iron is coupled to high-affinity carrier proteins (hemoglobin, ferritin, transferrin, and lactoferrin), which maintain a level of free iron exceedingly too low to sustain bacterial growth. Microbial pathogens adapt to such environment by synthesis and secretion of low-molecular-mass Fe^{3+}-chelating compounds, which can solubilize the metal bound to host binding proteins and transport it to the bacteria *via* a specific outer membrane receptor (Carniel, 2001).

1.3.2.1.1 Genetic structure of HPI

HPI, a typical GEI varying in sizes between 35 kb in *Y. pestis* to 43 kb in *Y. enterocolitica*, is inserted at an *asn* tRNA-locus. It is bordered by 18 bp-long direct repeats representing the core part of the hybrid attachment sites attL and attR. HPI can be divided into thee region: the mobility module, coding the P4 phage-like integrase; the core region carrying the yersiniabactin synthesis and transport genes; and a variable AT-rich region (Figure 1.3.7).

Yersinia pestis / pseudotuberculosis
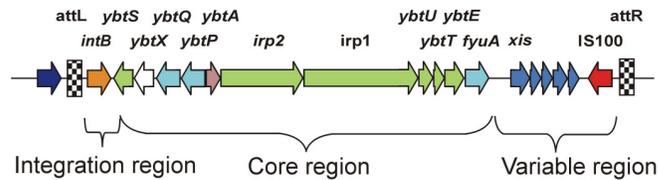
Figure 1.3.7. Genetic structure of the high pathogenecity islands from *Yersiniae* and *E. coli*. Asn tRNA genes are colored navy-blue, integrases – orange, yersiniabactin synthesis genes – green, genes involved in transport - blue, regulatory genes – pink, insertion elements – red, ORFs from variable region – dark blue, conjugal pilus assembly genes – magenta, plasmid maintenance genes – cyan, ORFs of unidentified function – white, attachment sites - checked.

The *Y. pseudotuberculosis* HPI can be associated with any of the three copies of the *asn* tRNA locus (Buchrieser et al., 1999), while the HPIs of *Y. enterocolitica* and *E. coli* are inserted specifically into the *asnT*. Locus HPI of *Y. pestis* is placed in yet another copy of the *asn* tRNA locus, which is associated with the so-called *pgm* (pigmentation) locus. Such localization may indicate a necessary linkage of two virulence traits, since the *pgm* locus is responsible for the blockage of the flea gut that is indispensable for transmission of bacteria to the warm-blooded host.

Six HPI-genes: *irp1*, *irp2*, *ybtU*, *ybtT*, *ybtE*, *ybtS* are involved in yersiniabactin synthesis, which is catalyzed by a multidomain, non-ribosomal peptide synthetase. Further genes, like *fyuA* - an outermembrane receptor, *ybtP* and *Q* – inner membrane

permeases, mediate the yersiniabactin transport into the cell. Expression of the biosynthetic and transport genes is activated by the transcriptional activator *ybtA*.

The AT-rich region shows significant diversification among species. In *Y. enterolitica* it contains several repeated sequences and is therefore responsible for the larger size of this island. In the *Y. pestis / Y. pseudotuberculosis* group it carries the excisionase gene and an untypical insertion element IS100. In most *E. coli* strains the AT-rich region together with the attR site, is deleted leading to immobilization of the island.

1.3.2.1.2 HPI transfer

HPI encodes a bacteriophage-related mobilization system consisting of the integrase and attachment sites: attL (which can be viewed as BOP' according to the phage attachment nomenclature) and attR (equivalent to POB'). Although never detected in a circular form in nature, the mobilization system of HPI is active since a module consisting exclusively of the integrase and synthetic attP site (created by substitution of the BO bacterial part of the left junction, with the PO part of the right junction) is capable of performing efficient excision and integration (Rakin et. al, 2001). When inserted into the asn tRNA gene, the endogenous integrase's promotor is replaced by the host asn tRNA promoter, which gives the integrated HPI the ability to synchronize with the physiological state of the host bacterial cell. This also implies that the activity of the HPI integrase upon integration and excision is dependant on two different promoters.

HPI encodes a phage P4-related integrase gene, which can produce a functional 420 aminoacid-long protein in all islands except for *Y. enterocolitica* where a premature introduction of the Stop codon results in a truncated peptide of 138

amino acids. The HPI integrase belongs to the tyrosine family of integrases and consists of three domains: the arm-type binding, core-binding, and catalytical.

In *Y. pestis* the loss of the HPI occurs at high frequencies (2×10^{-3}). This excision is, however, part of the deletion of the whole *pgm* locus mediated by homologous recombination between the two IS100 copies that flank the unstable region. The high deletion frequency of the pigmentation locus may mask however the lower frequency precise excision of the HPI alone, which is feasible due to presence of a functional integrase and complete att sites. In *Y. pseudotuberculosis* a site-specific excision is observed with an intermediate frequency of about 10^{-4}. The excision between the attL and attR sites corresponds to the mechanism of the P4-phage-mediated site-specific recombination. HPI of *Y. enterocolitica* is immobilized through both the unfunctional integrase and degenerated DR in attL. Stabilization of HPI in *E. coli* is even more advanced, with truncation of a large segment of the right border of the island, including the attR site.

Despite encoding the bacteriophage-related mobilization module, it is unlikely for HPI to be transported through transduction due to its large size and lack of phage packaging sequences. It should be noted that phage-associated recombination enzymes are detected also on other mobile elements like conjugative transposons or integrative and conjugative elements. On the other hand, much data speak in favor of the conjugation as the HPI transfer mechanism. HPI itself does not represent an integrated plasmid because it lacks genes needed for autonomous replication. It is, however, able to cointegrate with a shuttle vector if only that vector possesses a suitable attachment site, as shown by Antonenka et al. (2005) on a model of a promiscuous plasmid RP4 equipped with the artificial asn tRNA attachment site. HPI in a plasmid-conjugated form can be transferred to another bacterium where it

dissociates and reintegrates into the host chromosome by site-specific recombination. This hypothesis is further supported by discovery of a unique *E. coli* HPI (EcoR31), which, in contrast to the all known high pathogenecity island of this species, encodes a functional integrase system. Additionally, its AT-rich region contains systems for the pilus assembly, mating pair formation and DNA-processing typical for conjugative plasmids (Schubert et. al, 2004). This region is located exactly at the position of the IS100 insertion in *Y. pestis* and *Y. pseudotuberculosis*, which suggests a possible mechanism of its deletion.

1.3.2.2 Horizontally Acquired Island of Erwinia carotovora

In addition to animal pathogens, *Enterobacteriaceae* includes important plant pathogens, such as the *Erwinia carotovora* subsp. *atroseptica* (*Eca*), the first plant-pathogenic enterobacterium to be sequenced. *E. carotovora* is a commercially one of the world's most economically important pathogen restricted to potato in temperate regions, where it causes blackleg in the field and soft rot of tubers after harvest.

The *Eca* genome sequence reveals that genes responsible for basic cellular functions are shared with both the animal-pathogenic enterobacteria and nonenterobacterial genomes, and are located on the chromosome backbone. By contrast, genes that encode for proteins associated with the cell surface, pathogenicity, plant-associated functions, transport, and regulation are proportionally overrepresented in the horizontally acquired islands (HAIs) (Toth et al., 2006).

The *Eca* genome contains 17 major putative horizontally acquired islands. Two of the HAIs (HAIs 9 and 17) represent complete prophages, while the others contain phage remnants and genes for DNA mobilizing proteins associated with HGT. Seven HAIs encode integrases genes (HAIs 2, 4, 7, 11, 12, 13, and 16) and three of them, namely integrases located in HAI7, 12 and 13, are recognized as

phage P4-related. HAIs carry genes involved in wide range of processes such as: toxin and antibiotic biosynthesis (HAI2 and 10), nitrogen fixation (HAI14), adhesion (HAIs 2, 7, 8, and 15), and polysaccharide biosynthesis (HAIs 1 and 5).

HAI7 and HAI13 resemble HPI by the presence of the functional integrase genes together with some features of conjugated plasmids. Additionally, HAI7 encodes proteins responsible for the arsenate resistance along with a type IV secretion system, which was identified as a virulence determinant (Bell et al. 2004).

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 *S*trains and plasmids

<u>*E. coli* cloning strains</u>

One Shot TOP10	Invitrogen (Holland)
NovaBlue GigaSingles	Novagen (Canada)

<u>*E. coli* protein expression strains</u>

BL21 Star	Invitrogen (Holland)
BL21 Star(DE3)	Invitrogen (Holland)
BL21 Star(DE3) pLysS	Invitrogen (Holland)
BL21-CodonPlus(DE3)-RP	Stratagene (USA)
BL21-CodonPlus(DE3)-RIL	Stratagene (USA)
ArcticExpress(DE3)	Stratagene (USA)
B834 (DE3)	Novagen (Canada)

<u>Cyanobacteria strains</u>

Mastigocladus laminosus PCC7603	Institute Pasteur (France)
Synechococcus sp. PCC7002	Institute Pasteur (France)
Anabaena sp. PCC7120	Institute Pasteur (France)

<u>Plasmids</u>

pET 21a(+)	Novagen (Canada)
pET 30b(+)	Novagen (Canada)
pET 40b(+)	Novagen (Canada)
pET 46 Ek/LIC	Novagen (Canada)
pPICZα	Promega (USA)

2.1.2 Cell growth media and stocks

<u>Media</u>

For 1 liter LB medium: 10 g bacto tryptone

                                       5 g bacto yeast extract

                                       10 g NaCl

pH was adjusted to 7.0. For the preparation of agar plates the medium was supplemented with 15 g agar.

For 1 liter TB medium   12 g bacto tryptone

                                       24 g bacto yeast extract

                                       10 g NaCl

                                       4 ml glycerol

                                       900 ml deionized water

The medium was autoclaved, cooled; 100 ml sterile K phosphate and glucose were added. The final concentration of glucose was 0.5%.

For 1 liter K-phosphate, pH 7.1:   23.1 g KH_2PO_4

                                         125.4 g K_2HPO_4

For 1 liter of 2YT medium   16 g bacto tryptone

                                       10 g bacto yeast extract

                                       5 g NaCl

Minimal medium (MM) for uniform enrichment with ^{15}N:

For 1 liter MM: 0.5 g NaCl

                                       1.3 ml trace elements solution

                                       1 g citric acid monohydrate

                                       36 mg ferrous citrate

                                       4.02 g KH_2PO_4

                                       7.82 g K_2HPO_4 x $3H_2O$

                                       1 ml Zn-EDTA solution

                                       1 g $^{15}NH_4Cl$

pH was adjusted to 7.0 with NaOH, the mixture was autoclaved, and upon cooling the separately sterilized stock solutions were added in the following amounts:

25 ml glucose

560 µl thiamin

antibiotics

2 ml $MgSO_4$

Defined medium for selenomethionine labeling of proteins

For 1 liter of medium:

400 mg Ala, Gln, Glu, Arg, Gly

255 mg Asp,

125mg selenomethionine

125 mg cytosine, guanosine, uracil

100 mg Asn, Leu, His, Lys, Pro, Thr

100 mg Tyr

400 mg Ile, Val

50 mg Phe, thymine, thymidine

1.6 g Ser

10 mg $CaCl_2$

2 g NaAcetate

10 g K_2HPO_4

1 g citric acid

1.3 ml trace element solution

36 mg ferrous citrate

1 ml Zn-EDTA solution

1g NH_4Cl

pH was adjusted to 7.0 with NaOH, the mixture was autoclaved. Separately sterilized stock solutions were added to the cooled medium in the following amounts:

25 ml glucose

560 µl thiamine

antibiotics

2 ml $MgSO_4$

50 mg Cys, Trp, nicotinic acid

0.1 mg biotin

The second portion (125 mg) of the selenomethionine was added at the time of induction.

Pichia pastoris expression

For 1 liter of low-salt LB medium: 10 g bacto tryptone

                                                              5 g bacto yeast extract

                                                              5 g NaCl

                                                              pH 7.5

For 1 liter of YPD medium: 20 g bacto tryptone

                                                              10 g bacto yeast extract

After autoclaving 100 ml of 20% glucose (sterilized separately) was added. For the preparation of agar plates the medium was supplemented with 20 g agar.

For 1 liter of YPDS medium   20 g bacto tryptone

                                                              10 g bacto yeast extract

                                                              1 M sorbitol

After autoclaving 100 ml of 20% glucose (sterilized separately) was added. For the preparation of agar plates the medium was supplemented with 20 g agar.

Buffered glycerol-complex medium   10 g bacto yeast extract

                                                              20 g bacto tryptone

Yeast extract and tryptone were dissolved in 80o ml of deionized H_2O, autoclaved and cooled. Afterwards, following reagents (sterilized separately) were added:

                                                              100 ml glycerol

                                                              100 ml 1 M K phosphate pH 6.0

                                                              13,4 g Yeast Nitrogen Base

                                                              0.4 mg biotin

Buffered minimal methanol medium

Following reagents were added to 800 ml of autoclaved and cooled deionized H_2O:

                                                              100 ml methanol

                                                              100 ml 1 M K phosphate pH 6.0

                                                              13,4 g Yeast Nitrogen Base

                                                              0.4 mg biotin

<u>Stock solutions</u>

Ampicillin: 100 mg/ml of ampicillin in deionized H_2O, sterilized by filtration, stored in aliquots at -20°C until used. Working concentration: 100 µg/ml.

Chloramphenicol was dissolved in ethanol (0.34 g/10 ml) to the end concentration of 34 mg/ml. Working concentration: 34 µg/ml.

Kanamycin: 100 mg/ml of kanamycin in deionized H_2O, sterile filtrated and stored in aliquots at -20°C until used. Working concentration: 100 µg/ml.

Streptamycin 100 mg/ml of streptamycin in deionized H_2O, sterile filtrated and stored in aliquots at -20°C until used.

Zeocin: 100 mg/ml in deionized, autoclaved H_2O, supplied by manufacturer. For selection of *E. coli* 25 µg/ml working concentration was used, while *Pichia pastoris* selection required concentration 100 µg/ml.

IPTG: A sterile filtered 1 M stock of IPTG in distilled water was prepared and stored in aliquots at -20°C until used.

Glucose: 20% (w/v) in deionized H_2O, autoclaved.

Thiamin, 1%, in deionized H_2O, sterilized by filtration.

$MgSO_4$, 1 M, in deionized H_2O, sterilized by filtration.

Zn-EDTA solution:	5 mg/ml EDTA
	8.4 mg/ml $Zn(Ac)_2$

Trace elements solution:	2.5 g/l H_3BO_3
	2.0 g/l $CoCl_2$ x H_2O
	1.13 g/l $CuCl_2$ x H_2O
	9.8 g/l $MnCl_2$ x $2H_2O$
	2.0 g/l Na_2MoO_4 x $2H_2O$
	pH lowered with HCl

2.1.3 Protein purification buffers

<u>Buffers for immobilized metal-chelate chromatography (IMAC) in native conditions</u>

Lysis buffer	50 mM NaH_2PO_4
	300 mM NaCl
	10 mM imidazole
	pH 8.0

Wash buffer	50 mM NaH_2PO_4
	300 mM NaCl
	20 mM imidazole
	pH 8.0

Elution buffer	50 mM NaH_2PO_4
	300 mM NaCl
	250 mM imidazole
	pH 8.0

Buffers for IMAC in denaturing conditions

Binding buffer	6 M guanidinium chloride
	100 mM NaH_2PO_4 x H_2O
	10 mM Tris
	10 mM β-mercaptoethanol
	pH 8.0

Wash buffer	6 M guanidinium chloride
	100 mM NaH_2PO_4 x H_2O
	10 mM Tris
	10 mM β-mercaptoethanol
	pH 6.5

Elution buffer	6 M guanidinium chloride
	100 mM NaAcetate x $3H_2O$
	pH 3.0

Refolding buffer	200 mM arginine HCl
	1 mM EDTA
	100 mM Tris
	2 mM red GSH
	2 mM ox GSH
	10% (v/v) glycerol
	0.05% NaN_3
	pH 8.4

Protease buffers

Factor Xa cleavage buffer

50 mM Tris

100 mM NaCl

4 mM $CaCl_2$

0.05% NaN_3

pH 8.0

Enterokinase cleavage buffer

20 mM Tris

100 mM NaCl

2 mM $CaCl_2$

0.05% NaN_3

pH 7.5

Trypsin cleavage buffer

20 mM Tris

100 mM NaCl

0.05% NaN3

pH 7.6

Gel filtration chromatography buffers

Crystallization buffer

5 mM Tris

50 mM NaCl

10 mM β-mercaptoethanol

0.05% NaN_3

pH 7.5

PBS

140 mM NaCl

2.7 mM KCl

10 mM Na_2HPO_4

1.8 mM KH_2PO_4

0.05% NaN_3

pH 7.4

2.1.4 Buffer for DNA agarose gel electrophoresis

50X TAE buffer (for 1 l)

40 mM Tris-Acetate	242 g of Tris base
1 mM EDTA	100 ml of 0.5 M EDTA (pH 8.0)
Glacial acetic acid	57.1 ml

2.1.5 Reagents and buffers for the SDS-PAGE

Buffers and reagents

Anode buffer (+):   200 mM Tris pH 8.9

Cathode buffer (-):   100 mM Tris pH 8.25

100 mM tricine

0.1% SDS

Separation buffer: 1 M Tris pH 8.8

0.3% SDS

Stacking buffer: 1 M Tris pH 6.8

0.3% SDS

Separation acrylamide: 48% acrylamide

1.5% bis-acrylamide

Stacking acrylamide: 30% acrylamide

0.8% bis-acrylamide

Pouring polyacrylamide gels

Separation gel: 1.675 ml H_2O

2.5 ml separation buffer

2.5 ml separation acrylamide

0.8 ml glycerol

25 µl APS

	2.5 µl TEMED
Intermediate gel:	1.725 ml H_2O
	1.25 ml separation buffer
	0.75 ml separation acrylamide
	12.5 µl APS
	1.25 µl TEMED
Stacking gel:	2.575 ml H_2O
	0.475 ml stacking buffer
	0.625 ml stacking acrylamide
	12.5 µl 0.5 M EDTA, pH 8.0
	37.5 µl APS
	1.9 µl TEMED
Sample Buffer	0.225 M Tris
	50% glycerol
	5% SDS
	0.005% Bromophenol Blue
	0.25 M DTT
	pH 6.8

Protein visualization

Coomassie-blue solution:	45% ethanol
	10% acetic acid
Destaining solution:	5% ethanol
	10% acetic acid

2.1.6 Reagents and buffers for western blots

Transfer buffer	20 mM CAPS
	10% methanol

Wash buffer   10 mM Tris

                                                      150 mM NaCl

                                                      0.05% Tween20

                                                      pH 8.0

Alkaline phosphatase buffer   100 mM Tris

                                                      200 mM NaCl

                                                      50 mM $MgCl_2$

                                                      pH 9.5

Primary antibody solution   1:2000 diluted in the wash buffer

Secondary antibody (AP-linked)   1:2000 diluted in wash buffer

Substrate for alkaline phosphatase (AP)   BCIP (Sigma); 1 tablet dissolved in

                                           10 ml of water

2.1.7 Enzymes and other proteins

Kpn I   New England BioLabs (USA)

Not I   New England BioLabs (USA)

Sal I   New England BioLabs (USA)

Pme I   New England BioLabs (USA)

Dpn I   Stratagene (USA)

CIP   New England BioLabs (USA)

Pfu turbo DNA Polymerase   Stratagene (USA)

Phusion HF DNA Polymerase   Finnzymes (Finland)

T4 DNA Ligase   New England BioLabs (USA)

Xa Factor   Novagen (Canada)

Enterokinase   Novagen (Canada)

Trypsin   GIBCO (Hollland)

Mouse monoclonal Anti-His antibodies   Santa Cruz Biotech (USA)

Goat anti-mouse antibodies   Santa Cruz Biotech (USA)

BSA   New England BioLabs (USA)

2.1.8 Kits and reagents

QIAquick PCR Purification Kit	Qiagen (Germany)
QIAprep Spin Miniprep Kit	Qiagen (Germany)
QIAGEN Plasmid Midi Kit	Qiagen (Germany)
QIAGEN Plasmid Maxi Kit	Qiagen (Germany)
Rapid Ligation Kit	Roche (Germany)
Complete Protease Inhibitor Cocktail	Roche (Germany)
pET LIC cloning Kits	Novagen (Canada)
Pre-Crystallization Test (PCT)	Hampton Research (USA)
FoldIt Screen	Hampton Research (USA)

2.1.9 Protein and nucleic acids markers

Prestained Protein Marker	New England BioLabs (USA)
100 BP DNA marker	New England BioLabs (USA)
1Kb DNA marker	New England BioLabs (USA)

2.1.10 Chromatography equipment, columns and media

ÄKTA explorer 10	Amersham Pharmacia (Sweden)
Peristaltic pump P-1	Amersham Pharmacia (Sweden)
Fraction collector RediFrac	Amersham Pharmacia (Sweden)
Recorder REC-1	Amersham Pharmacia (Sweden)
UV flow through detector UV-1	Amersham Pharmacia (Sweden)
BioloLogic LP System	Biorad (USA)
HiLoad 26/60 Superdex S75pg	Amersham Pharmacia (Sweden)
HiLoad 16/60 Superdex S75pg	Amersham Pharmacia (Sweden)
HiLoad 16/60 Superdex S200pg	Amersham Pharmacia (Sweden)
HiLoad 10/30 Superdex S75pg	Amersham Pharmacia (Sweden)
HiLoad 10/30 Superdex S200pg	Amersham Pharmacia (Sweden
NiNTA-agarose	Qiagen (Germany)

2.2 Methods

2.2.1 General remarks on constructs' design

The choice of the constructs' amino and carboxyl termini may decide about protein solubility, stability, and later success or failure of crystallization attempts. Disrupting secondary structure element can interrupt folding and render an inclusion bodies-directed protein. Alternatively, flexible fragments (like remnants of neighboring domains present due to erratic construct design) can be subjected to spontaneous cleavage, contributing to protein instability. Furthermore, flexible parts of protein are known to inhibit crystallization or be responsible for obtaining crystals of low quality, useless for structure determination. Studies on domain organization employing limited proteolysis, protein sequencing, mass spectrometry and NMR spectroscopy (Rehm et al., 2002), can help in determination of approximate boundaries of the proteins' domains.

2.2.2 Choice of the expression system

E. coli, as the simplest and most popular expression system offers several advantages like high yield of target protein, fast growth rate and low costs. It naturally suited for expression of prokaryotic proteins like cyanobacterial biliproteins or integrases from *Enterobacteriacea.*

For efficient protein expression, a tightly controlled, yet still promoting high-level transcription and translation, systems are preferred, such as a pET system based on the phage T7 RNA polymerase. In the pET system, the protein coding sequence of interest is cloned downstream of the T7 promoter (not recognized by host polymerases) and protein expression is achieved by IPTG induction of a chromosomally integrated T7 RNA polymerase expressed from the *lacUV5* promoter

The system provides several levels of control due to additional plasmids, which express T7 lysosyme, a natural T7 RNA polymerase inhibitor enabling a high level of control over a basal (i.e.uninduced) expression of transgenes.

Bacterial cells differ substantially from eukaryotic ones, which influences expression efficiency of the proteins derived from the higher organisms. The frequencies of codon usage in *E. coli* genes are distinct from that of the Eucaryota. As a result, a pool of amino acyl-tRNAs and aa-tRNA synthetases present in *E. coli* also differs from the one found in human cells. In particular, Arg codons AGA, AGG, CGG, CGA, Ile codon AUA, Leu codon CUA, Gly codon GGA and Pro codon CCC are rarely used. Therefore, various genes that contain these codons may be inefficiently expressed by and result in premature termination of the synthesized protein or misincorporation of amino acids. This impact appears to be highest when these codons are present near the N-terminus or in clusters. It was shown that the yield of a protein, whose genes contain rare codons, can be dramatically improved when the cognate tRNA is increased within the host (Rosenberg et al., 1993). *E. coli* strains with artificially introduced copies of genes encoding the tRNAs, either at the genomic level or in a plasmid, are commercially available (Rossetta from Invitrogen; BL21 RP or RIL from Stratagene).

Bacterial cells are unable to perform many of the post-translational modifications performed by higher eukaryotic cells such as proteolytic processing, folding, disulfide bond formation and glycosylation. Therefore, eukaryotic proteins tend to be expressed with much poorer results, often ending up as inactive inclusion bodies. Utilization of eukaryotic systems like mammalian cells expression or the baculovirus system solves these problems, but is combined with high costs and low

efficiency. As a yeast, *Pichia pastoris* combines easiness of manipulation and culture with ability of the post-translational modifications.

2.2.3 DNA techniques

2.2.3.1 Preparation of plasmid DNA

The isolation of plasmid DNA from *E. coli* was carried out using dedicated plasmid purification kits from Qiagen. The kit employs a standard alkalic lysis of the precipitated bacteria in the presence of RNAse and strong ionic detergent, SDS, followed by neutralization/DNA renaturation with acetate. For purification, a crude cell lysate was loaded onto a silica gel column, washed with an ethanol-containing buffer, and eluted in a small volume, yielding up to 20 μg of the plasmid DNA.

2.2.3.2 PCR

A polymerase chain reaction was employed to amplify desired DNA fragments and genes, introduce restriction sites, Stop codons and sequences encoding restriction and protease cleavage sites. The primers were prepared according to standardized principles regarding the length, GC-content, melting temperature and occurrence of secondary structures of the hairpin type. Primers used for cloning and mutageneses are listed in the tables in respective parts of the thesis.

Two kinds of recombinant thermostable DNA polymerases were used, each operating at slightly different conditions:

	Melting temp.	Annealing temp.	Synthesis temp.
Phusion HF	98°C	primer melting - 5°C	72 °C
Pfu Turbo	95°C	primer melting - 5°C	68 °C

Synthesis time was adjusted to the length of the sequence to be amplified, calculating 1 kb / min for Pfu Turbo polymerase and 1 kb/ 15 s for Phusion HF polymerase. Synthesis step in the last cycle was extended to 10 min to assure completion of the reaction.

2.2.3.3 Digestion with restriction enzymes

Usually, 1-2 units of each restriction enzyme were used per 1 µg of plasmid DNA to be digested. The digestion was performed in a buffer specified by the manufacturer at the optimal temperature (37 °C) for 5-16 h. Because fragments' ends that occurred after digestion were cohesive, 5'-ends of the vector were dephosphorylated using calf intestine phosphatase (CIP) to eliminate the possibility of plasmid recirculation (possible when double-digestion does not occur with 100% efficiency). CIP treatment was performed with 1 unit of enzyme per 3 µg of plasmid DNA, at 37 °C for 1 h.

2.2.3.4 Purification of PCR and restriction digestion products

DNA obtained from restriction digestion, CIP treatment or PCR was purified from primers, nucleotides, enzymes, buffering substances, salts, agarose, SYBR-Gold, and other impurities, using a silica-gel column (QIAquick PCR Purification Kit, Qiagen). The QIAquick system uses a simple bind-wash-elute procedure. A binding buffer was added directly to the PCR sample or other enzymatic reaction, and the mixture was applied to the spin column. Nucleic acids were adsorbed to the silica-gel membrane in the high-salt conditions provided by the buffer, while impurities and short fragments of single- or double-stranded DNA were washed away. Pure DNA was eluted with a small volume of 10 mM Tris pH 8.0 or water.

2.2.3.5 Site directed mutagenesis

Site directed mutagenesis of was performed by means of PCR. The mutagenic oligonucleotide primers were designed according to several rules: they possessed 25-45 bases in length, their melting temperatures were greater than 78 °C, they had a GC content of more than 40%, terminated in C or G bases, and the desired mutation was in the middle of the primer with ~10-15 bases of a correct sequence on both. Vector containing a copy of a gene of interest was used as a DNA template.

Following temperature cycling, the product was treated with *Dpn* I (10 U, 37 °C, for 2 h). The *Dpn* I endonuclease (target sequence: 5´-Gm6ATC-3´) is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. 2 µl of the mixture were used to transform XL1-Blue or Top10 chemically competent cells. Plasmid DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen) and was subjected to verification by automated sequencing.

2.2.3.6 Agarose gel electrophoresis of DNA

For verification of the presence and length of PCR or restriction digestion products, agarose gel electrophoresis was performed. For this purpose 1% agarose in TBE buffer plus Sybr Gold (Invitrogen) was prepared. The solution was poured into a horizontal gel chamber to cool down and polymerise. The DNA samples were mixed with 6x sample buffer prior loading. Electrophoresis was carried out at 100-120 V DC. Results were evaluated using UV illumination.

2.2.4 Transformation of competent cells

2.2.4.1 Chemical transformation of E. coli

A number of factors were shown to increase efficiency of transformation. Such factors include: prolonged incubation of bacteria with CaCl2, addition of multiple cations, such as Mg^{2+}, Cs^{2+} or Rb^{2+} and treatment of bacteria with dimethyl sulfoxide (DMSO), polyethylene glyol, hexaminecobalt, and dithiothreitol in the presence of both monovalent and divalent cations (Chung et al., 1989).

E.coli cells rendered chemically competent by rinsing bacteria harvested during log-growth phase with 100 mM $MgCl_2$, followed by resuspension in 100 mM $CaCl_2$ supplemented with 15% glycerol.

3 µl of a ligation mix or ca. 50 µg of plasmid DNA was added to 50 µl of chemically competent cells. The mixture was incubated on ice for 30 min followed by a heat shock of 30 s at 42 °C, short cooling on ice, and addition of 250 µl of glucose- and magnesium-containing medium. Heat shock aids in recovery and bacteria survival by inducing the heat shock genes. The cells were than incubated for 1 h at 37°C without selective pressure for sufficient time to be given for expression of antibiotic resistance genes. Plating on LB agar plates that included selective antibiotic enabled recovery of those cells that actually received the DNA.

2.2.4.2 Transformation by electroporation of Pichia pastoris cells

Pichia pastrois cells were grown in YPD medium at 30°C until OD_{600} 1,3-1,5. Cells were harvested and washed twice with ice-cold, sterile H_2O. After last washing step, cells were resuspended in sterile 1 M sorbitol and used immediately for electroporation. Transformation by electroporation employs high voltage to temporarily puncture bacterial cell walls. High mortality of cells treated this way

implicates the need for high concentrations of bacteria. Preparation of electrocompetent cells was resolved by thorough washing of pelleted cells in order to remove salt present in growth media, which would otherwise cause short-circuiting and mass destruction of the cells.

5-10 μg of plasmid DNA (linearized by Pme I and purified by PCR Purification Kit, Qiagen) was added to 1.5 ml of prepared cells in electroporation cuvette and incubated on ice for 5 min. The electroporation was performed in an Gene pulser vessel at 1100 V. Then the suspension was mixed with 1 ml of 1 M sorbitol and transferred to 15 ml sterile tube. After 30 min incubation at 30°C (without shaking) cells were spread on YPD agar plates containing Zeocin.

2.2.5 Protein chemistry methods & techniques

2.2.5.1 Protein expression

Competent cells of the suitable bacteria strain were transformed according to protocol described above. The overnight cultures of 20 ml were used as an inoculum for the 1 liter culture containing suitable antibiotic. Typically, the cells were induced at OD_{600} = 0.7-0.8 by addition of IPTG (1 mM final concentration) and cells were grown for the next 3 h with vigorous shaking at 37°C, or overnight in 20°C.

The cells were harvested by centrifugation (5500 x G, 20 min, 4 °C) and resuspended in PBS (when meant for the denaturing conditions protocol), or lysis buffer, (when meant for the native conditions prep).

2.2.5.2 Sonication

Sonication is a simple method used for the disruption of cells by ultrasounds. The high frequency sound wave induces formation of cavitations in the solution, the collapse of which produces sheer forces able to break cells. The advantage of this

method for cell disruption over French press, freezing and thawing or lysozyme digestion is comparatively low viscosity of the lysate due to the nucleic acid fragmentation. Pulsed mode of operation was applied (output control 8, 60% duty cycle) and sonication was carried out on ice, in 3 steps of 5 min each, with 5 min intervals between steps, to avoid overheating of the sample.

2.2.5.3 Protein purification

Purification protocol differed substantially between denaturing and native conditions. In case of denaturing conditions, pellet obtained after centrifugation (20000xg, 30 min) of sonificated cells was solubilized in binding buffer containing guanidinium chloride overnight in 37°C. Afterwards suspension was centrifuged to remove plasma membranes and cell debris. The resulting supernatant was incubated with a Ni-NTA slurry (Qiagen), equilibrated previously with the same buffer for 1 h at room temperature with gentle agitation. Next, the mixture was loaded onto an empty column and washed with wash buffer. The protein was eluted by application of buffer with low pH. The fractions containing the desirable protein were pooled and concentrated. Subsequently, the refolding of the protein was performed. For this purpose, the protein sample was stepwise diluted in suitable refolding buffer in a 1:50 volume proportion. The refolding mixture was left with stirring at 4°C. After 3 days, the mixture was concentrated, dialyzed against an appropriate buffer, and subjected to gel filtration.

In case of native conditions, supernatant obtained after centrifugation (20000xg, 30 min) was mixed with Ni-NTA resin preequilibrated with lysis buffer and incubated at least for 2 h at 4°C with gentle agitation. Afterwards, mixture was loaded onto an empty column and washed with wash buffer. The protein was eluted by application of buffer with high concentration of imidazole (elution buffer). The

fractions containing the desirable protein were pooled, concentrated and subjected to gel filtration on Superdex 75 (protein separation range 250 – 20 kDa) or Superdex 200 (protein separation range 75 – 5 kDa). If protein fused with the tag was acceptable for further procedures, gel filtration was carried out in a buffer suitable for intended purpose (PBS for NMR measurements, crystallization buffer for crystallization screening and optimization).

When removal of that was necessary, the gel filtration served as a step of buffer exchange the one suitable for the given protease. Cleavage was performed by enterokinase (cleavage of tags), factor Xa (cleavage at the introduced sites inside the protein) or trypsin (limited proteolysis). Proteolysis was monitored by SDS-PAGE electrophoresis, and when completed, was stopped by addition of PMSF (serine protease inhibitor). The proteins of interest were separated from the cleaved fragments and protease by second round of gel filtration. In case of tag removal, gel filtration was preceded by short incubation with NiNTA resin, which resulted in retention of the protein with still uncleaved tag.

2.2.5.4 SDS polyacrylamide gel electrophoresis (SDS PAGE)

SDS polyacrylamide gel electrophoresis was performed at various stages of purification to verify samples' purity, identity of the eluted proteins in chromatography purification, progression of enzymatic digestion etc. Due to the small size of the expressed proteins, tricine gels were applied (Schagger and von Jagow, 1987). Tricine, used as the trailing ion instead of glycine, allows a resolution of small proteins at lower acrylamide concentrations than in glycine-SDS-PAGE systems. A resolution of proteins, especially in the range between 5 and 20 kDa, is thus achieved without the necessity to use urea.

The guanidinium HCl-free protein samples were prepared by mixing 20 µl of protein solution with 5 µl of sample buffer (SB) followed by 3 min incubation at 100 °C.

Due to rapid precipitation of SDS in contact with guanidine, the samples to be examined by PAGE after Ni-NTA chromatography under denaturing conditions had to be prepared in a following fashion: 20 µl of the protein solution in a denaturing buffer was diluted with 400 µl 20% trichloroacetic acid (TCA). The sample was incubated for 5 min at room temperature followed by centrifugation for 5 min at 20 000 x g. Supernatant was discarded by suction, and the precipitated protein pellet was washed by vortexing with 400 µl ethanol. After centrifugation and ethanol removal, the protein pellet was resuspended in 20 µl of 2x SB and the sample was boiled for 3 min.

2.2.5.5 Visualization of separated proteins

For visualization of the protein bands, the gels were stained in a Coomassie-blue solution. Background was cleared by incubation of the gel in a destaining solution. Both processes were greatly accelerated by brief heating with microwaves of the gel submerged in an appropriate solution.

2.2.5.6 Electroblotting

Electroblotting technique was applied as a primary step of N-terminal sequencing of band of interest, to identify the protein of interest in the mixtures of proteins, or, alternatively, to confirm a tag removal.

Briefly, the unstained polyacrylamide gel after protein separation, was applied on top of nitrocellulose in case of Western blot applications or PVDF membrane for N-terminal sequencing. PVDF membranes were activated with methanol prior usage. The gel and the membrane were surrounded with a stack of pre-wetted Whatmann papers and electrotransfer was performed for 1 h with constant voltage 100 V.

Membranes designated for N-terminal sequencing were stained according to standard procedure applicable to polyacrylamide gels with restriction to special purity of used reagent (sterile-filter prior usage). Membranes were dried and chosen bands were subjected to Edman sequencing after excision.

Blots meant for immunodetection were blocked in 5% milk (Fluka) in wash buffer for 1 h at 37°C, or overnight at 4°C. Solution of antibodies were applied sequentially, separated by excessive washing with the wash buffer. Each antibody was incubated for 1.5 – 2 h in RT. Finally, the membrane was rinsed with the alkaline phosphatase buffer. Blot was incubated in the substrate solution until desired clarity of image was obtained.

2.2.5.7 Determination of protein concentration

The concentration of proteins in solution was estimated by means of the Bradford colorimetric assay. 5 µl of the protein sample was added to 1 ml (10 x diluted stock) Bradford reagent (BioRad) into a plastic cuvette. After gentle mixing, A_{595} was measured and converted to the protein concentration on the basis of a calibration curve prepared for known concentrations of BSA. A precise protein concentration determination was performed spectrophotometrically. Absorption at 280 nm was measured and converted to a protein concentration on the basis of theoretical extinction coefficients. It has been shown that it is possible to estimate the molar extinction coefficient Eλ(Prot) of a protein from knowledge of its amino acid composition (Gill and Hippel, 1989). From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine residues do not absorb appreciably at wavelenghts >260 nm, while cystine does) at a given wavelength λ the extinction coefficient of a protein can be computed using the equation:

Eλ(Prot) = Numb(Tyr)xExtλ(Tyr) + Numb(Trp)xExtλ(Trp) + Numb(cystine)xExtλ(cystine)

The absorbance (A, optical density) can be calculated using the following formula: Aλ (Prot) = Eλ(Prot) / Molecular weight.

2.2.6 NMR spectroscopy

All NMR experiments were carried out at 300 K on a Bruker DRX 600 spectrometer equipped with a triple resonance, triple gradient 5 mm probehead. The samples contained typically 0.1-0.5 mM protein in the PBS buffer supplemented with 10% ^2H_2O. All 1D ^1H NMR spectra were recorded with a time domain of 32 K complex points and a sweep-width of 10,000 Hz. The 2D ^1H-^{15}N-HSQC spectra were recorded with a time domain of 1K complex data points with 128 complex increments with a sweep width of 8 kHz in the ^1H dimension and 2 kHz in the ^{15}N dimension.

2.2.7. Crystallization

Crystallization trials were performed using protein purified at the last step by the gel filtration (to assure homogenecity of the sample in terms of polymerization). Crystallization buffer typically contained 50 mM NaCl, 5 mM Tris pH 7.5, and in cases of proteins containing disulphide bridges, 10 mM β-mercaptoethanol. For proteins especially prone for precipitation (i.e. PecE and some of the PecF constructs) ionic strength of the buffer was raised to 100 – 150 mM NaCl.

Sitting drop vapor technique in 96-well format was used for screening. A typical screen was composed of the following: Index (Hampton Research) Classics, PEGs, MPDs, Anions, Cations, pHClear I, pH Clear II (Qiagen), and Wizard I and II (Emerald Biosystems) at 4°C and 20°C. Optimization of the most promising conditions was performed in both, sitting drop and hanging drop techniques, through manipulation of crystallization reagent composition, drop size and protein:reagent ratio.

3. Results and Discussion

3.1. Nicastrin

3.1.1 Construct design and cloning

The nicastrin gene was fished out from the human cDNA library. A construct comprising residues 1-636 (omitting the 33 aminoacid-long signal sequence) was prepared and cloned into the pET46 LIC vector (Novagen) according to the manufacturer's instructions.

Table 3.1.1 List of primers used in the work

No.	Sequence	Name
1	GACGACGACAAGATGGCTACGGCAGGGGGT	hNCT_fishout_For
2	GAGGAGAAGCCCGGT*TTA*CTCAAGCTCTTTGCTGGCGAT	hNCT_Rev
3	GACGACGACAAGATGAACTCAGTGGAGAGG	hNCT-signal_For
4	GACGACGACAAGATGGACGACAGGGTTG	hNCT_239-476_For
5	GAGGAGAAGCCCGGT*TTA*GTCGCTGAAGTT	hNCT_239-476_Rev
6	GACGACGACAAGATGAGCGCAGTGG	hNCT_264-378_For
7	GAGGAGAAGCCCGGT*TTA*GACAGCAGGG	hNCT_264-378_Rev
8	GAGGAGAAGCCCGGT*TTA*GCCCAGCGAATT	hNCT_1/4_Rev
9	GACGACGACAAGATGAATGGTTTGGCTTATG	hNCT_2/4_For
10	GAGGAGAAGCCCGGT*TTA*CCCTTGAAAGAA	hNCT_2/4_Rev
11	GACGACGACAAGATGGAAACTTTTGACTAC	hNCT_3/4_For
12	GACGACGACAAGATGACAGTTCAGGCT	hNCT_4/4_For
13	CC**GTCGAC**AACTCAGTGGAGAGGAAG	hNCT_Sal_For
14	CT**GCGGCCGC**CTCAAGCTCTTTGCTGG	hNCT_Not_Rev
15	GGAAGAGGTACCATG**AACTCAGTG**GAGAGGAAG	hNCT_Kpn_PICZ_For
16	CCTTT**GCGGCCGC***TTA*CTCAAGCTCTTTGCT	hNCT_Not_PICZ_Rev

For-forward, **Rev**-reverse, restriction sites in **bold**; stop codons in *italics*

Table 3.1.2. List of constructs produced in the work

No.	Name	Vector	Primers used
1	Full-length	pET46	3,2
2	hNCT (239-476)	pET46	4,5
3	hNCT (263-378)	pET46	6,7
4	hNCT (1-476)	pET46	3,6
5	hNCT (136-666)	pET46	7,2
6	hNCT (1-299)	pET46	3,10
7	hNCT (136-476)	pET46	10,5
8	hNCT (300-666)	pET46	11,2
9	hNCT (1-135)	pET46	3,8
10	hNCT (136-299)	pET46	9,10
11	hNCT (300-476)	pET46	11,5
12	hNCT (477-666)	pET46	12,2
13	Full-length	pET40b	13,14
14	Full-length	pPICZα	15,16

3.1.2. Optimization of growth and expression conditions

The nicastrin protein was obtained only in insoluble (inclusion bodies) fractions. Furthermore, the expression in *Escherichia coli* BL21(DE3) cells resulted in production of the protein attenuated at different steps of translation (and hence of different lengths, Figure 3.1.1A), probably due to presence of rare codons. To solve this problem, several bacteria strains were tested as expression hosts, namely: CodonPlus-RP and CodonPlus-RIL (suitable for the expression of GC-rich and AT-rich genes, respectively) as well as in the in-house prepared CodonPlus-RIL cells supplemented with the pRARE plasmid (containing tRNA for further seven rare codons). Usage of the latter strain enabled the production of the full-length nicastrin protein (Figure 3.1.1B).
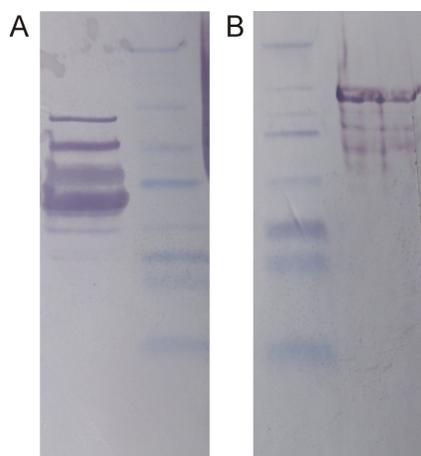
Figure 3.1.1. Western blot of the nicastrin ectodomain expressed in standard (**A**) compared to the optimized (**B**) conditions.

Extensive efforts were made to obtain a soluble protein. Lowering the temperature (growth temperature 20°C) after induction is one of most commonly used strategies. This approach was used individually and in combination with the ArcticExpress cells (Stratagene) coexpressing the cold-adapted chaperonins from the Antarctic isolate, *Oleispira Antarctica,* that are induced by lowering the temperature to 10°C.

Several media other than standard LB were tested i.e. the 2YT medium and the Terrific Broth medium. The idea behind utilization of highly enriched media is to promote expression of minute amounts of protein if they were present in a soluble fraction. None of the tested media and conditions produced the protein in the native fraction. Western blots of the inclusion bodies fractions showed, however, that the expression in Terrific Broth at 20°C overnight gives the most homogenous protein preparation, and therefore this approach was used in further studies (Figure 3.1.2).
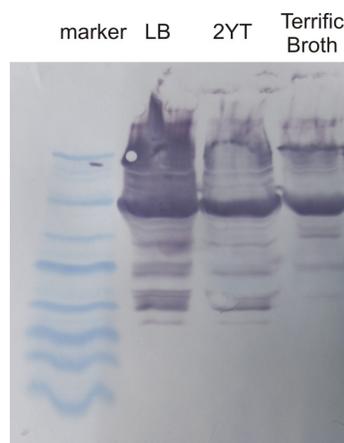
Figure 3.1.2. Western blot image of expression of the nicastrin ectodomain in different media.

3.1.3. Refolding

Due to difficulties in obtaining nicastrin in a native state several refolding trials were performed. Optimizing the correct refolding conditions remains an empirical process. The pH of the refolding buffer can be an important factor, since proteins tend to have a characteristic pH range at which they can fold efficiently and reach their active state. Usable refolding additives span wide range of small molecules like aminoacids, sugars, polyols, and salts. Additives fall into two categories: refolding enhancers and aggregation supressors. First group encompasses certain aminoacids, like glycine and alanine, as well as polyols, which increase protein-protein interactions and stabilize target protein causing it to collapse into a compact structure. Other additives, like arginine, proline, mild detergents, and polyethylene glycol, act as aggregation suppressors *via* binding to the hydrophobic regions of the folding intermediate. Presence of an adequate redox system has been shown to be crucial for functional refolding because the correct disulfide bond formation tends to be the limiting step during the folding process. The use of a suitable redox system, consisting of oxidized and reduced thiol reagents, allows disulfide reshuffling, i.e.

wrong disulfide bonds will be reduced again because they are not protected by the correct structural context. One of the most popular redox couples is the oxidized and reduced glutathione (GSSG/GSH). "Smaller" redox couples like cystein/cystine can also be useful in some cases.

To perform the refolding screen, the nicastrin ectodomain was purified from inclusion bodies. Inclusion bodies, obtained as pellet after sonication and centrifugation of disrupted cells, were solubilized in the 6M guanidinium chloride through vigorous stirring overnight. The solution was cleared by centrifugation and applied to NiNTA resin under reducing conditions. Binding was conducted for 2 h with agitation. Afterwards, slurry was applied on the column and washed with guanidinium chloride of pH lower than the pH for optimal binding but not low enough to elute the Histagged protein. Elution was performed stepwise by guanidinium chloride solution of pH 3.0, deprived of reducing agents. pH of 3.0 prevents the creation of disulphide bridges through the sulphydryl group protonation.

A standard refolding buffer, which contained 200 mM arginine, 10% glycerol and the GSSG/GSH redox system, was used as a primary option. Refolding was performed by a rapid (pulse) dilution technique. The volume of the refolding buffer was adjusted to the amount of protein, so that the final concentration was about 0.1 mg/ml. The volume of the injected protein in the guanidinium chloride solution was kept low, not to let the concentration of denaturant exceed 0.15 M. This approach failed to produce a soluble protein already at the stage of refolding itself.

In order to establish optimal conditions of refolding for nicastrin, the FoldIt Screen (Hampton Research) was used. The screen consists of 16 conditions spanning different ionic strengths, pH, target protein concentrations and refolding

additives (arginine, guaninidinium chloride, PEG3350, sucrose, lauryl maltoside, EDTA or bivalent ions).

Given amounts of the protein (expressed and purified as described above) were injected stepwise into refolding conditions, taking care of keeping a constant end concentration of guanidinium chloride (0.3 M) and refolding was performed for 24 h at 4°C with gentle agitation. Afterwards, the samples were centrifuged and dialyzed to buffers that did not contain any additives but had the salt concentration and pH of the original refolding conditions kept. After dialysis the protein remained soluble in 2 samples, i.e. FoldIt Screen conditions 8 and 16 (Figure 3.1.3).

Composition of the most successful conditions is as follows: FoldIt Screen 8: 55 mM MES pH 6.5, 264 mM NaCl, 11 mM KCl, 11 MM EDTA, 0.055% w/v PEG3350, 550 mM L-arginine monohydrate, 1 mM GSH / 0.1 mM GSSG, with the final protein concentration 1 mg/ml; FoldIt Screen 16: 55 mM Tris pH 8.2, 264 mM NaCl, 11 mM KCl, 2.2 mM MgCl$_2$, 2.2 mM CaCl$_2$, 0.055% w/v PEG 3350, 550 mM guanidinium chloride, 440 mM sucrose, 550 mM L-arginine monohydrate, 1 mM GSH / 0.1 mM GSSG, with the final protein concentration 1 mg/ml.
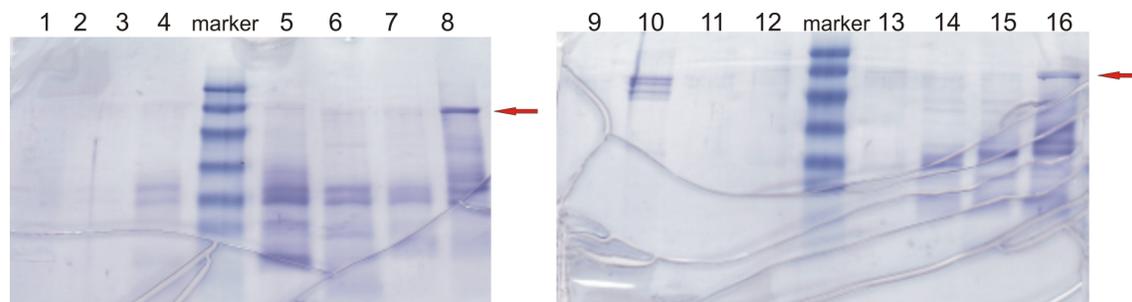
Figure 3.1.3. SDS-PAGE of nicastrin refolded in the FoldIt Screen conditions (1-16). The red arrow indicates position of the nicastrin band.

The samples were subjected to gel filtration to establish their aggregation status (Figure 3.1.4).
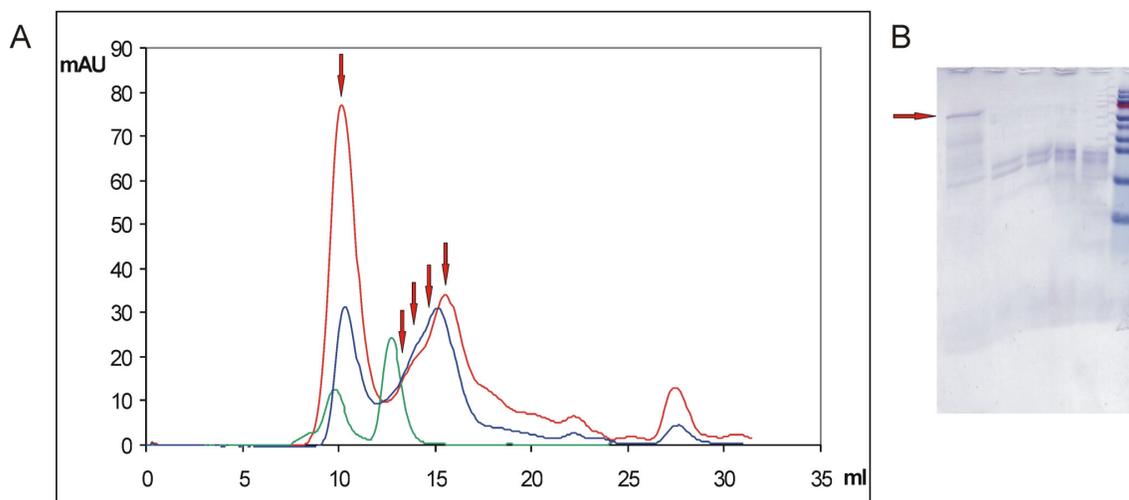
Figure 3.1.4. (**A**) Gel filtration of nicastrin after refolding in FoldIt Screen buffers 8 (blue line) and 16 (red line). The green line indicates standards of masses 67 kDa and 25 kDa given in a suitable scale. Red arrows indicate the fractions selected for electrophoresis from refolding in FoldIT Screen condition 16; (**B**) SDS-PAGE image of chosen fractions. Lines 1-5 - selected fractions, line 6 - protein marker. The arrow indicates the protein band of the desired mass.

Since both conditions produced the protein which could be characterized as soluble and unaggregated, a large scale purification and refolding was performed following the initial protocol, with some changes due to scaling-up introduced, namely lowering the end concentration of nicastrin in the refolding buffer to 0.1-0.2 mg/ml and increasing the concentration of GSH and GSSG to 2 mM. After 3 days of refolding, followed by dialysis, samples were subjected to size exclusion chromatography (Figure 3.1.5). The protein remained soluble, not aggregated (as confirmed by gel filtration) but turned out to be unfolded as revealed by 1D-NMR.
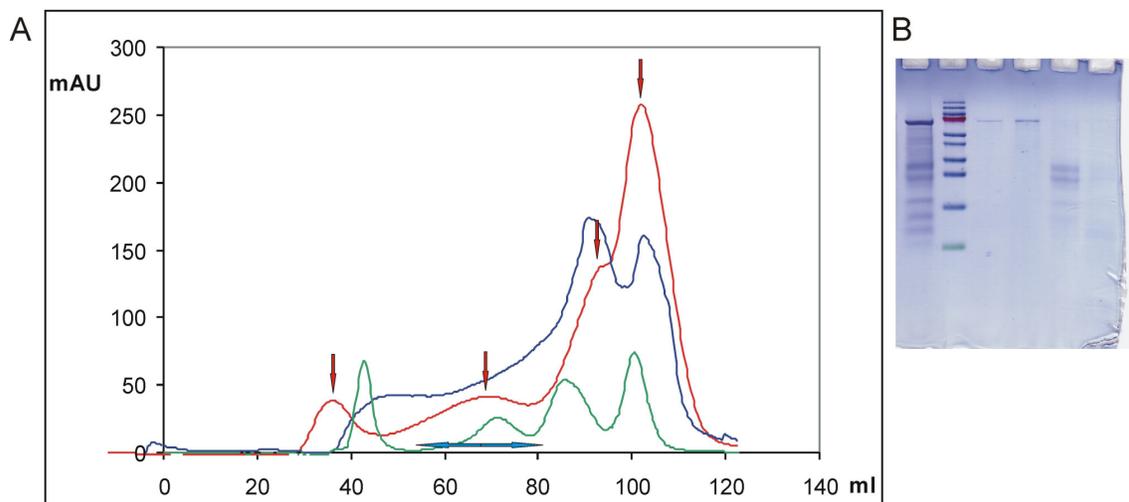
Figure 3.1.5. (A) Gel filtration of nicastrin after refolding in FoldIt Screen buffers 8 (blue line) and 16 (red line). The green line indicates standards of masses 666, 256, 67 and 25 kDa given in suitable scale. Red arrows indicate fractions from FoldIt Screen 16 – the refolded protein selected for electrophoresis, blue arrows indicate the pooled area; **(B)** SDS-PAGE image of chosen fractions. Line 1 - hNCT at the final stage of refolding, line 2 - protein marker, lines 3-6 - selected fractions.

3.1.4 The peptidase M20-like domain

Complex architecture of the protein domains is common in the human genome and other similarly advanced organisms. Computational biology defines conserved domains based on recurring sequence patterns or motifs. When working with protein of unknown structures, some hints about the existence and localization of domains may be obtained through homology searches. This approach can lead to characterization of substructures that can have ability to fold independently since domains can be thought as distinct functional and / or structural units of a protein. These two classifications coincide rather often and what is found as an independently folding unit of a polypeptide chain also carries specific function.

To identify conserved domains in a protein sequence, the CD-Search service uses the reverse position-specific BLAST algorithm. The query sequence is

compared to a position-specific score matrix prepared from the underlying conserved domain alignment. Inspection of the human nicastrin sequence with the Conserved Domain Search engine (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) revealed presence of a domain belonging to the peptidase M20/M25/M40 superfamily, spanning residues 241-466 of nicastrin (Figure 3.1.6). This family includes a range of zinc metallopeptidases that belongs to several families in the peptidase classification. Members of family M20 are glutamate carboxypeptidases, while the M25 family contains X-His dipeptidases. The presence of this domain is consistent with previous knowledge about presumed evolutionary connections of nicastrin (Fagan et al., 2001).

```
          CD Length: 227   Bit Score: 244.76   E-value: 4e-65
            10        20        30        40        50        60        70        80
     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
241  RVVVAATRLDSRSFFWNVAPGAESAVASFVTQLAAAEALQKA-PDVTTLPRNVMFVFFQGETFDYIGSSRMVYDMEKGKF 319

  1  KVVLVTARMDSTSMFDGVSLGAMSSLSGFIVLLAAADALSKAlPDISNLKRNVLFAFFNGESYDYIGSQRFVYDMENGKF 80

            90       100       110       120       130       140       150       160
     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
320  P--------VQLENVDSFVELGQVALRTSLELWMHTDpvSQKNESVRNQVEDLLATLEKSGAGVPAVILRRPNQSQPLPP 391
 81  PsdrththpISPDNIDYMLEIGQVGKATSRKFYLHVD--AARNQSVKTQTLDLLDRIEKSLRSGNFKVLPASTSNPGLPP 158

           170       180       190       200       210       220       230
     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*.
392  SSLQRFLRAR-NISGVVLADHSGAFHNKYYQSIYDTAENINvsypewlSPEEDLNFVTDTAKALADVATVLGRALY 466
159  SSLQSFLRANpNFSAVVLADRPTEFENRFYHSILDDAENIN-------SDTEDLNEKDSQQMSVVNAASLVARALY 227
```

Figure 3.1.6. The peptidase family M20/M25/M40. The best-scoring hit on this query sequence is by the member pfam05450.

A construct comprising the detected domain was designed taking into account the helix and β-sheet distribution as identified through the secondary structure prediction tool (Consensus Secondary Structure Prediction, Pôle Bioinformatique Lyonnaisse) Alignment of sequences of nicastrin and the best-scoring hit is presented in Figure 3.1.11. During the design of the construct, the presence of large, hydrophobic residues on either of the ends was avoided. Therefore, the construct comprises residues 239 to 476. A C-terminal prolongation of the construct, compared

to the peptidase dimerization domain, was added because of the detection of a large, partially hydrophobic helix spanning residues 450-470.

Nicastrin (residues 239-476) was expressed in conditions established for the full-length protein, and could not be obtained in a soluble form. Refolding, using the FoldIt Screen was performed as described before. After concentration and clarification of refolding factorials, the protein was present in significantly larger number of conditions, namely conditions no.: 1, 4, 6, 7, 8, 10, 12, 13, 15 and 16 (Figure 3.1.7).
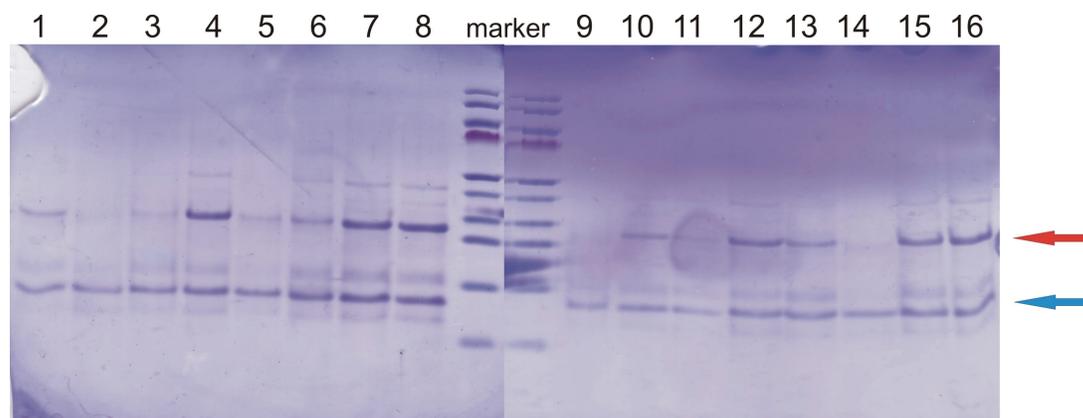
Figure 3.1.7. SDS-PAGE of the nicastrin 239-476 construct refolded in the FoldIt Screen conditions 1-16. The red arrow indicates position of nicastrin construct band; blue arrow indicates position of degradation product

Conditions were divided into four groups based on the original pH (6.5 *versus* 8.2) and salt concentrations (264 Mm NaCl, 11 mM KCl *versus* 10.64 mM NaCl, 0.44 mM KCl). After dialysis, the protein was present in solution in the following conditions: 4, 7, 10, 13, 15, and 16 (Figure 3.1.8A). The presence of the Histag was confirmed by the Western blot (Figure 3.1.8B).

A

B
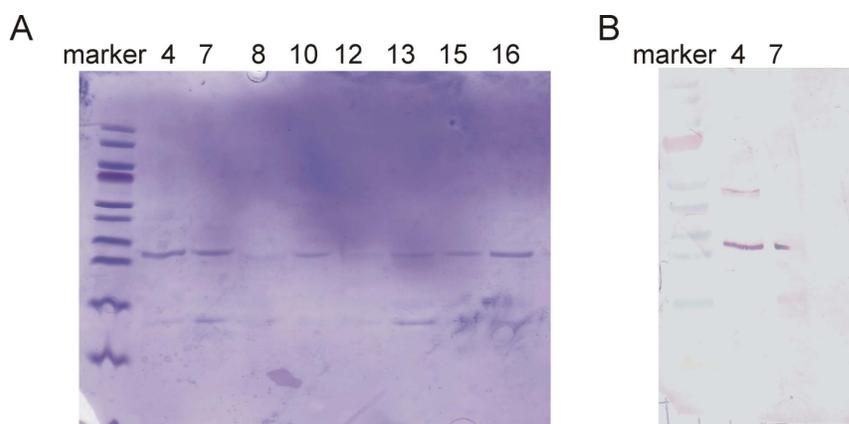
Figure 3.1.8. (**A**) SDS-PAGE of the nicastrin 239-476 construct refolded in subsequent FoldIt Screen conditions and dialysed against a suitable buffer. (**B**) Western blot image of chosen (4 and 7) conditions.

The most abundant protein fractions obtained after refolding were subjected to gel filtration. The most successful trials are presented in Figure 3.1.9.
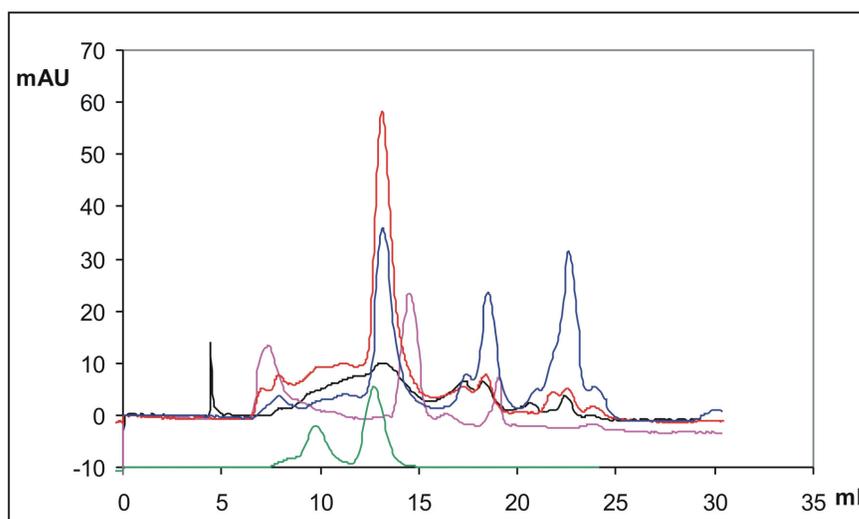
Figure 3.1.9. Gel filtration of nicastrin (239-476) after refolding in the FoldIt Screen buffers 4 (black line), 7 (pink line), 8 (blue line) and 16 (red line). The green line indicates standards of masses 67 kDa and 25 kDa.

After refolding a separate band of an approximately 12 kDa mass was identified (Figure 3.1.7, blue arrow). This fragment was present in elution profiles from most of gel filtrations (Figure 3.1.9, the smaller peak at approximately 18 ml).

This band could not be visualized in the Western blot indicating that it was not Histagged. Based on the mass spectrometry data, this fragment was identified as spanning residues 264-378 and was cloned as an independent construct. This construct, covering residues 264-378, was tested for expression under native conditions but failed to produce a soluble protein. Under denaturing conditions large amounts of the protein were obtained. After standard NiNTA purification, a refolding trial was carried out in 100 mM Tris buffer, pH 8.4, supplemented with 200 mM arginine and 10% glycerol as additives. The protein remained soluble but, as shown by gel filtration, it was mostly aggregated (Figure 3.1.10). 1D-NMR of the end product of refolding showed it to be unfolded.
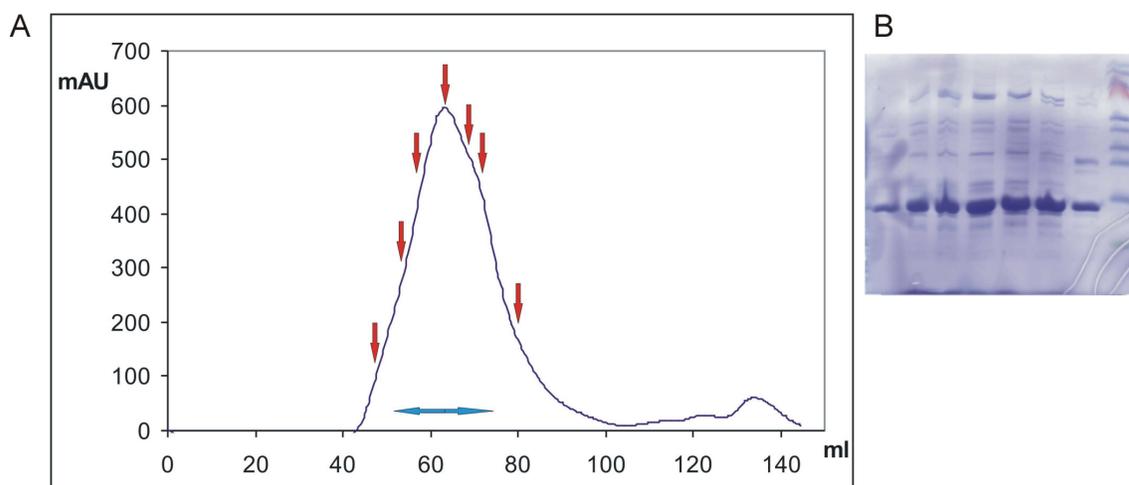
Figure 3.1.10. (**A**) Gel filtration of the refolded nicastrin construct 264-378. Red arrows indicate fractions chosen for electrophoresis, blue arrows indicate the pooled fractions, (**B**) SDS-PAGE image. Lines 1-7 – selected fractions, line 8 – protein marker.

3.1.5 Sequence-scan for soluble fragments

The aminoacid sequence of nicastrin was systematically scanned throughout its entire length for fragments encompassing approximately three fourths, halves or quarters of the full-length protein to find fragments that could be expressed in a

native fraction. Boundaries of constructs were placed in regions predicted as unstructured according to the secondary structure prediction tool (Figure 3.1.11).

Primers used in preparation of constructs are shown in Table 3.1.2, while constructs' residue coverage, together with basic biophysical parameters, are grouped in Table 3.1.3 (constructs 4-12). All proteins were expressed under conditions established for the full-length protein. They were obtained in moderate to large amounts, however, only as inclusion bodies.

Table 3.1.3. Biophysical properties of the investigated proteins

No.	Construct	No. of aa	MW (Da)*	E_{280} (M^{-1} cm^{-1})**	pI
1	hNCT (1-636) (fl)	651	72813	1.290	5.61
2	hNCT (264-378)	130	14484	0.585	5.05
3	hNCT (239-476)	253	28111	0.958	5.05
4	hNCT (1-476)	491	54497	0.988	5.39
5	hNCT (136-636)	516	58000	1.316	5.37
6	hNCT (1-300)	314	34817	0.946	5.88
7	hNCT (136-476)	356	39685	0.936	5.06
8	hNCT (301-636)	352	39858	1.530	5.49
9	hNCT (1-135)	150	16675	1.032	6.34
10	hNCT (136-300)	179	20005	0.786	5.46
11	hNCT (301-476)	192	21542	0.995	5.08
12	hNCT (477-636)	175	20178	1.960	6.44

*Mol. weight is calculated for the proteins with the N-terminal Histag MAHHHHHHVDDDDKM and for reduced cysteines.

**E(Prot) = Numb(Tyr) x Ext(Tyr) + Numb(Trp) x Ext(Trp) + Numb(Cystine) x Ext(Cystine), where "Ext" stands for the extinction coefficient at 280 nm.

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
Nicastrin NSVERKIYIPLNKTAP-VRLLNATHQIG-QSSISGDTGVIHVVEKEEDLQWVLTDGPNPPYMVLLESKHF
DSC       ------EEE---------------EEEEE------EEEEEE-------EEE-------EEEEEE--HH
MLRC      --HHHEEEH-------EEEEEE--------------EEEEEEE----EEEEEE------EEEEE-HHHH
PHD       ----EEEEE-------EEE----------------EEEEEEEE-----HHHH------EEEEE---HH
Sec.Cons. ----?EEEE--------EEE----------------EEEEEEE-----?EEE-------EEEEE---HH

                   80        90       100       110       120       130       140
                    |         |         |         |         |         |         |
Nicastrin TRDLMEKLKGRTSRIAGLAVSLTKPSPASGFSPSVQ-PNDGFGVYSNSYGPEFAH-REIQWNSLGNGLAY
DSC       HHHHHHHH------HHEEEEEE------------------EEEE---------------------------
MLRC      HHHHHHHHHH---EEEEEEEE--------------------EE------------------------E
PHD       HHHHHHHHHH----EEEEEEE----------------------------------------------EE
Sec.Cons. HHHHHHHHHH----?EEEEEEE-------------------EE------------------------E

                  150       160       170       180       190       200       210
                    |         |         |         |         |         |         |
Nicastrin EDFSFPIFLLEDENETKVIKQ-YQDHNLSQNGSAPTFPL-AMQLFSHMHAVISTAT-MRRSSIQSTFSIN
DSC       ------EEE--HHHHHHHHHH--------------HHHHHHHHH------HHHHH-----------
MLRC      -----EEEEE----HHHHHHHHHHH-------------HHHHHHHHHHH------HHHH----------
PHD       E----EEEEEE---HHHHHHHHHHHH-------------EHHHHHHHH------HHH------------
Sec.Cons. -----EEEEE----HHHHHHHHHH-------------HHHHHHHHHH------HHHH----------

                  220       230       240       250       260       270       280
                    |         |         |         |         |         |         |
Nicastrin PEIV-DPLSDYNVWSMLKPINTTGTLKPDDRVVVAATRLDSRSFFWNVAPGAESAVASFVTQLAAAEALQ
DS-       ------------EEE---------------EEEEEE----HHHHH------HHHHHHHHHHH----
MLR-      ------------EHEH--------------EEEEEEE----EEEE------HHHHHHHHHHHHHHHHHH
PHD       --------------EE---------------EEEEEE----HHHHH------HHHHHHHHHHHHHHHHHH
Sec.Cons. ------------E?E?--------------EEEEEEE----HHHHH------HHHHHHHHHHHHHHHHH

                  290       300       310       320       330       340       350
                    |         |         |         |         |         |         |
Nicastrin KAPDVTTLPRNVMFVFFQGETFDYIGSSRMVYDMEKGKFPVQLENVDSFVELGQVALRTSLELWMHTDPV
DSC       HH---------EEEE------------EEE-------EEEE--HHHHHHHHHH------EEEE-----
MLRC      H---------EEEEEE----EEEE---EEEEH------HHHHHHHHHHHHHHHH-------
PHD       H---------EEEEEEE----------HHHHHHHHHH------E-----EEEEEEE--------------
Sec.Cons. H---------EEEEEE-----------EEE?H-------EEE---HHHHHHHHHH------??-------

                  360       370       380       390       400       410       420
                    |         |         |         |         |         |         |
Nicastrin SQKNESVRNQVEDLLATLEKSGAGVPAVILRRPNQSQPLPPSSLQRFLRARNISGVVLADHSGAFHNKYY
DSC       ------HHHHHHHHHHHH------EEEE------------HHHHHHH-----EEEE---------E
MLRC      ----HHHHHHHHHHHHHH-----EEEEE----------HHHHHHH----EEEEE----------HHH
PHD       --------HHHHHHHHHHHH------EE-----------HHHHHHH-----EEEE----------
Sec.Cons. ------HHHHHHHHHHHHHH----EEEE-----------HHHHHH-----EEEE---------?

                  430       440       450       460       470       480       490
                    |         |         |         |         |         |         |
Nicastrin QSIYDTAENINVSYPEWLSPEEDLNFVTDTAKALADVATVLGRALYELAGGTNFSDTVQADPQTVTRLLY
DSC       EE-------EEEE----------------HHHHHHHHHHHHHHHHHH--------------HHHHHHH
MLRC      HHHH------------------HHHHHHHHHHHHHHHHHHHHHH------------HHHHHHHH
PHD       E------------------------HHHHHHHHHHHHHHHHHHH----------HHHHHHHH
Sec.Cons. E?------------------------HHHHHHHHHHHHHHHHHH-----------HHHHHHHH

                  500       510       520       530       540       550       560
                    |         |         |         |         |         |         |
Nicastrin GFLIKANNSWFQSILRQDLRSYLGDGPLQHYIAVSSPTNTTYVVQYALANLTGTVVNLTREQ-QDPSKVP
DSC       HHHH-----HHHHH--------------EEEE-------EEEEEHHHH--------------------
MLRC      HHHHH----HHHHHHHHHHHH------EEEEEE-----EEEEEEHHHH----EEEE------------
PHD       HHHHH---HHH-------------EEEEEE------EEEHHHHHHH-HHHHHHHHHH--------
Sec.Cons. HHHHH----HHHHH-------------EEEEEE-----EEEEEHHHH----????-----------

                  570       580       590       600       610       620       630
                    |         |         |         |         |         |         |
Nicastrin SENKDLYEYSWVQGPLHSNETDRLPR-VRSTARLARALSPAFELSQWSSTEYSTWTESRWKDIRARIFLI
DSC       -----EEEEEEE-----------HHHHHHHHHHHH----------------------HHEEEHH
MLRC      -------E----------------HHHHHHHHHHHH-----EE------------HHHHHHHHHHHEEEE
PHD       ------------------------EE-EHHH-E-------EE----------HHHHHHHHHHHEEEE
Sec.Cons. -------E----------------HHHHHHHHHHHH------E-----------HHHHHHHHHHHEEEE

                  640       650       660       670
                    |         |         |         |
Nicastrin ASKELELITLTVGFGILIFSLIVTY-INAKADVLFIAPREPGAVSY
DSC       HHHHHHHH-----EEEEEE--EEEEEE-----EEEE----------
MLRC      ---HHHEHEHHHHHHHHHHHHHHHHHHH----EEEEEE---------
PHD       E---EEEEEE---HHHEE-EEEEEEE-----EEE----------
Sec.Cons. ?--HHHEHE??--?HHHEE-EEEEEE-----EEEE---------
```

Figure 3.1.11. Secondary structure prediction of nicastrin using the Consensus Secondary Structure Prediction Tool (Pôle Bioinformatique Lyonnaisse). Comparison among the solutions found by four different algorithms. E - β-sheet, H – helix.

3.1.6 Periplasmic expression

Targeting protein to extracellular compartments offers alternative to expression in cytoplasm, especially useful when the latter results in toxic or misfolded proteins. The periplasm compartment of Gram-negative bacteria provides more oxidative environment in comparison to that of the cytoplasm. It also contains many proteins facilitating folding (proline cis isomerases) and the disulphide bridge formation (DsbA) and isomerization (DsbC). Another advantage is that the periplasmic localization may protect the recombinant protein from action of proteases and facilitate its subsequent purification.

To test influence of the periplasmic localization on expression of human nicastrin, the sequence encompassing a full-length protein was cloned into the pET40b vector (Novagen). This vector is designed to produce a fusion protein consisting of the N-terminal DsbC (the disulphide bond isomerase) fused with target protein. Studies suggest that this particular member of the disulphide bond-formation system tends to be the most useful in production of recombinant protein (Kurokawa et al., 2001). DsbC provides not only the signal sequence which enables translocation of the fused construct to the periplasm in a highly soluble form, but also provides enzymatic activity which is overexpressed and localized in the proximity to the recombinant protein.

BL21(DE3) cells transformed with the pET40b nicastrin construct were grown as described before. The periplasm proteins were liberated from bacteria cells through the osmotic shock procedure. Briefly, harvested cells were dissolved in a hypertonic solution (20% sacharose, 30 mM Tris pH 8.0 and EDTA to final concentration of 60 mM), the mixture was stirred for 10 min and the cells were spinned down at 10000xg. This procedure was meant to induce the cytoplasm

shrinkage due to the hypertonic environment while addition of EDTA was supposed to increase the permeability of cell wall. Pretreated cells were incubated in ice-cold water to release periplasm proteins through exposing cells to the hypotonic solution. For sake of further NiNTA resin purification (which is sensitive to presence of EDTA) 5 mM $MgCl_2$ was added. The solution with the periplasm content was cleared by centrifugation and applied to the NiNTA resin. After examination of all obtained fractions no nicastrin was found in soluble fractions.

Reasons for failing in this approach might be multiple. Firstly, periplasmic expression tends to be lower than cytoplasmic. Another obstacles might have been the lack of simultaneous overexpression of DsbD (the activity modulator of DsbC) or environment more oxidative to that found in the natural nicastrin maturation compartment, i.e. the endoplasmic reticulum. Furthermore, a protein in order to pass the cytoplasmic membrane has to change its conformation. If this is incompatible with translocation, the process cannot be completed and the target protein is degraded.

3.1.7 *Pichia pastoris* expression

Due to the encountered difficulties in obtaining the correctly folded nicastrin and the fact that nicastrin ectodomain, being an extracellular protein, has to undergo necessary posttranslational modifications, an eukaryotic expression system was tested. The *Pichia pastoris* system combines benefits of the prokaryotic expression (like simplicity of culture, strong, chemically induced promoter, and possible high levels of recombinant proteins) with advantages of the expression in eukaryotic cells (i.e. protein processing, folding, and posttranslational modifications).

The nicastrin ectodomain gene was chosen to be expressed under a strong inducible AOX1 (alcohol oxidase) promoter. *Pichia pastoris* is capable of

metabolizing methanol as the sole carbon source. The first step in the metabolism of methanol is the oxidation of methanol to formaldehyde by the enzyme, alcohol oxidase. Expression of this enzyme is tightly regulated and induced by methanol to very high levels - typically >30% of the total soluble protein in the cells.

The nicastrin gene was cloned and propagated in the PICZα vector, which enables selection (with use of zeocin) in *Escherichia coli* as well as in *Pichia pastoris* cells. Recombinant proteins are meant to be expressed as fusions with the N-terminally located peptide, which encodes the *Saccharomyces cerevisiae* mating factor prepro signal sequence. This enables efficient secretion of most proteins from *Pichia pastoris*. The protein of interest is also fused to the C-terminally located c-myc epitope and Histag to facilitate detection and purification.

The *Pichia pastoris* system differs from bacterial expression systems. Autonomously replicating vectors are generally not used for expression and vectors are integrated into the genome of the host. Such integration events happen during transformation and because the organism has a dominant homologous recombination system, virtually all integration events occur between regions of significant homology between sequences in the vector and the host genome (e.g., between the *AOX1* promoter sequences). Transformation is conducted by electroporation using a linearized plasmid (in this case with the Pme I restrictase).

When establishing the recombinant protein production in *Pichia pastoris*, one has to test any optimize strain used in the expression. Three types host strains are available. First, the wild type with regard to the *AOX1* and *AOX2* gene types (e.g. GS115), which grows in methanol at normal rates (methanol utilization plus, or Mut$^+$ phenotype). Secondly, hosts with genes totally or partially deleted (e.g. KM71 with *AOX1* gene replaced with the *Saccharomyces. cerevisiae ARG4* gene) which have to

rely on much weaker *AOX2* gene for alcohol oxidase production and grow on methanol at a slow rate (methanol utilization slow, or Mut^S phenotype). Finally, some hosts have both AOX genes deleted and therefore are unable to grow in methanol (methanol utilization minus, or Mut$^-$ phenotype).

Strains GS115 and KM71 were transformed and tested for nicastrin expression. Briefly, clones of strain KM71H were cultivated in the buffered glycerol-complex medium for 24 h and subsequently, after washing, the medium was changed to the buffered minimal methanol medium. Cells were cultivated for further 6 days and supplemented with methanol every 24 h. Samples were taken daily and subjected to SDS-PAGE electrophoresis followed by Coomassie Blue staining and Western blot analysis with anti-Histag antibodies. In the case of GS115 strain, the procedure was identical, with exception of the different total culture time (4 days) and more frequent media sampling (every 12 h). None of tested clones showed the presence of the nicastrin ectodomain in the medium as well as in the intracellular fraction.

3.2 Phycocyanobilin Lyases

3.2.1. PecF

3.2.1.1 Initial constructs, expression, and purification conditions

The phycoerythorocyanin lyase subunit F (PecF) isolated from cyanobacterium *Mastigocladus laminosus* and cloned into the pET30a vector was a gift from Dr. S. Böhm. Typical purification protocol included expression in BL21(DE3) bacteria in the standard LB medium at 37°C for approximately 3 h. Afterwards, the bacteria were harvested and disrupted by sonication. The cleared solution was applied to the NiNTA column in the standard lysis buffer. After washing with the lysis buffer, supplemented with 20 mM imidazole, the Histagged proteins were eluted by 250 mM imidazole and subjected to gel filtration in a buffer suitable for crystallization or enzyme digestion. If a tag cleavage was carried out, the mixture of a cleaved protein, a tag, the enzyme, and possibly some remains of the uncleaved, tagged protein, were subjected to another round of gel filtration in order to ascertain the purity of the sample prior crystallization.

3.2.1.2. Construct design and crystallization

Several constructs of PecF were designed and screened for crystallization. Typical screen included: Qiagen's Classics, PEGs MPDs, Cations, Anions, pHClear 1 and 2 and Hampton Research Index) at 20ºC and 4ºC in the sitting drop 96-well plates. Concentrations used were established based on the evaluation of the precipitate in the PreCrystallization Test (Hampton Research) or on one of popular sparse matrix screens (Index or Classics).

Initial studies were based of the pET30a construct, containing a long (approximately 5 kDa) N-terminal Histag cleavable by enterokinase.

Taking advantage of the existence of a natural, internal "Histag" (residues His53-His56) that could enable easy purification, an untagged construct in pET19 was designed.

Since 1D-NMR studies of the full length PecF indicated some level of unfolding, a systematic screen for the shortened, yet still soluble protein was performed. The list of primers used in this work is shown in Table 3.2.1.

Table 3.2.1. List of primers used in the work

	No.	Sequence	Name
PecF	1	GACGACGACAAGATGTCCTCCGTCGCTACTGCTGCTGTTGAA	pecF_57-212_For
	2	GAGGAGAAGCCCGGT*TTA*ATCCGATTCGCCAGCTACAGCTGC	pecF_57-212_Rev
	3	GACGACGACAAGATGATCGCAGCTCGCTTTAGC	pecF_1-201_For
	4	GAGGAGAAGCCCGGT*TTA*TCGCGATCGCACAAC	pecF_1-201_Rev
	5	GAGGAGAAGCCCGGT*TTA*AGCTACAGCTGCAGA	pecF_1-191_Rev
	6	GAGGAGAAGCCCGGT*TTA*CTGCGGTGTGGCAAT	pecF_1-181_Rev
	7	GAGGAGAAGCCCGGT*TTA*GCGGTCAGTAAC	pecF_1-149_Rev
	8	GAGGAGAAGCCCGGT*TTA*GGCAATGCGGCG	pecF_1-130_Rev
	9	GAGGAGAAGCCCGGT*TTA*GGCAACTTCCGT	pecF_1-119_Rev
	10	GACGACGACAAGATGTTTACAAATAGCAAGC	pecF_6-109_For
	11	GAGGAGAAGCCCGGT*TTA*AAATGCTTTAGC	pecF_6-109_Rev
PecE	12	GACGACGACAAGATCCGTTATTACGCAGCTTGGTGGTTGGGA	pecE_27-218_For
	13	GAGGAGAAGCCCGGT*TTA*GTTAGGCACCTTTGCTGCTAAAAT	pecE_27-218_Rev
	14	TTAGCAGCAAAGGTGCCTAAC<u>ATCGAAGGTCGT</u>AGCTTAAAATTGCTGAATTTA	pecEx_For
	15	GTTTTAAATTCAGCAATTTTAAGCT<u>ACGACCTTCGAT</u>GTTAGGCACCTTTGCTGCTAAAATG	pecEx_Rev
	16	<u>ATCGAAGGT</u>CGTGCAGCTTGGTGGTTGGGAAAG	pecEx2_For
	17	<u>ACGACCTTCGAT</u>GTAATAACGGATTTGGTTATC	pecEx2_Rev
	18	GACGACGACAAGATGACTGCTGCTTGT	pecE_1-226_For
	19	GAGGAGAAGCCCGGT*TTA*TTTTAAATTCAG	pecE_1-226_Rev
	20	GAGGAGAAGCCCGGT*TTA*CAAATACCGCGC	pecE_1-167_Rev

CpcE/F	21	GACGACGACAAGATGACTAATGAACTAATTAAT	cpcF_7120_For
	22	GAGGAGAAGCCCGGT*TTA*TTGACTTTGAGC	cpcF_7120_Rev
	23	GACGACGACAAGATGATAGAACCC	cpcE_7120_For
	24	GAGGAGAAGCCCGGT*TTA*CAACAATGAATC	cpcE_7120_Rev
	25	GACGACGACAAGATGCGATCGCTAATTTC	cpcF_7002_For
	26	GAGGAGAAGCCCGGT*TTA*CGGAAGACTCAG	cpcF_7002_Rev
	27	GACGACGACAAGATGTCGGACTGGCAG	cpcE_7002_For
	28	GAGGAGAAGCCCGGT*TTA*GAGCAGACTATC	cpcE_7002_Rev

For-forward, Rev-reverse, stop codons in *italics*, protease cleavage sites <u>underlined</u>

According to the SMART (Simple Modular Architecture Research Tool, http://smart.embl-heidelberg.de), first 56 aminoacids were recognized as unstructured, followed by four E-Z type HEAT repeats, known to be present in the cyanobacterial phycocyanin lyases. Therefore, a PecF(57-212) construct was designed and expressed. Under a standard protocol this protein was produced exclusively in inclusion bodies. Further optimization of growth conditions (lowering the expression temperature, altering the medium composition by addition of ethanol or glycerol) also failed to produce the soluble protein.

Using this experience with shortening the N-terminus of the protein, constructs including various truncations of the C-terminus were designed. The starting points for the study were constructs covering residues 1-201, 1-191, and 1-181, since these proteins have been successfully studied in functional assays with PecF(1-201), maintaining partial enzymatic activity (Zhao et al. 2005).

Two longer proteins were stable and screened for crystallization, while PecF(1-181) underwent spontaneous degradation and precipitation. PecF(1-191) yielded two or three distinct bands (when uncleaved or cleaved with enterokinase, respectively) that could be separated from the original protein through gel filtration (Figure 3.2.1).
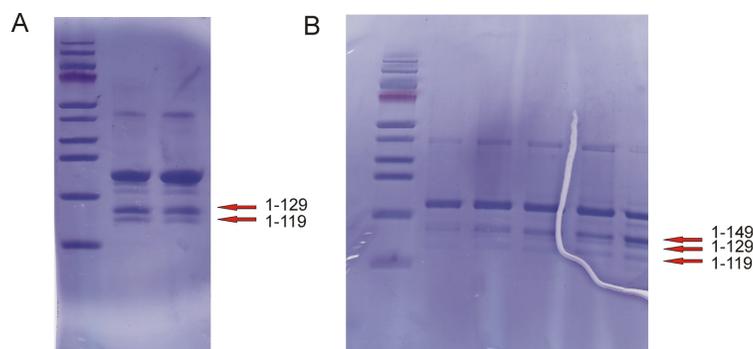
Figure 3.2.1. SDS-PAGE image of samples from PecF(1-191) with the Histag, (**A**), and cleaved with enterokinase (**B**). Arrows indicate the identified bands.

These fragments were identified through mass spectrometry as covering residues 1-149, 1-130 and 1-119, cloned as independent constructs, and screened for crystallization. Properties of the examined proteins are summarized in Table 3.2.2, while 1D-NMR spectra of the screened constructs are shown in Figure 3.2.3. It can be noticed that the overall folding of the protein improves along with the shortening of the construct.

Table 3.2.2

Construct	Vector	Primers	MW (kDa)	Screening	Cleavage of tag	
					Uncleaved	Cleaved
1-212	pET30a	-	27.4 (22.6)	+	+	+
1-212	pGEMEX	-	21.4	+	No tag	
57-212	pET46	1,2	16.7	- (insoluble)		
1-201	pET46	3,4	23.1 (21.4)	+	+	+
1-191	pET46	3,5	21.9 (19.9)	+	+	+
1-181	pET46	3,6	21.0	- (unstable)		
1-149	pET46	3,7	17.4	+	+	-
1-130	pET46	3,8	15.4	+	+	-
1-119	pET46	3,9	14.1	+	+	-
6-109	pET46	10,11	12.5 (10.8)	+	+	+

Molecular weight is calculated for the protein containing a tag. Values in brackets represent the molecular weight after removal of the tag.
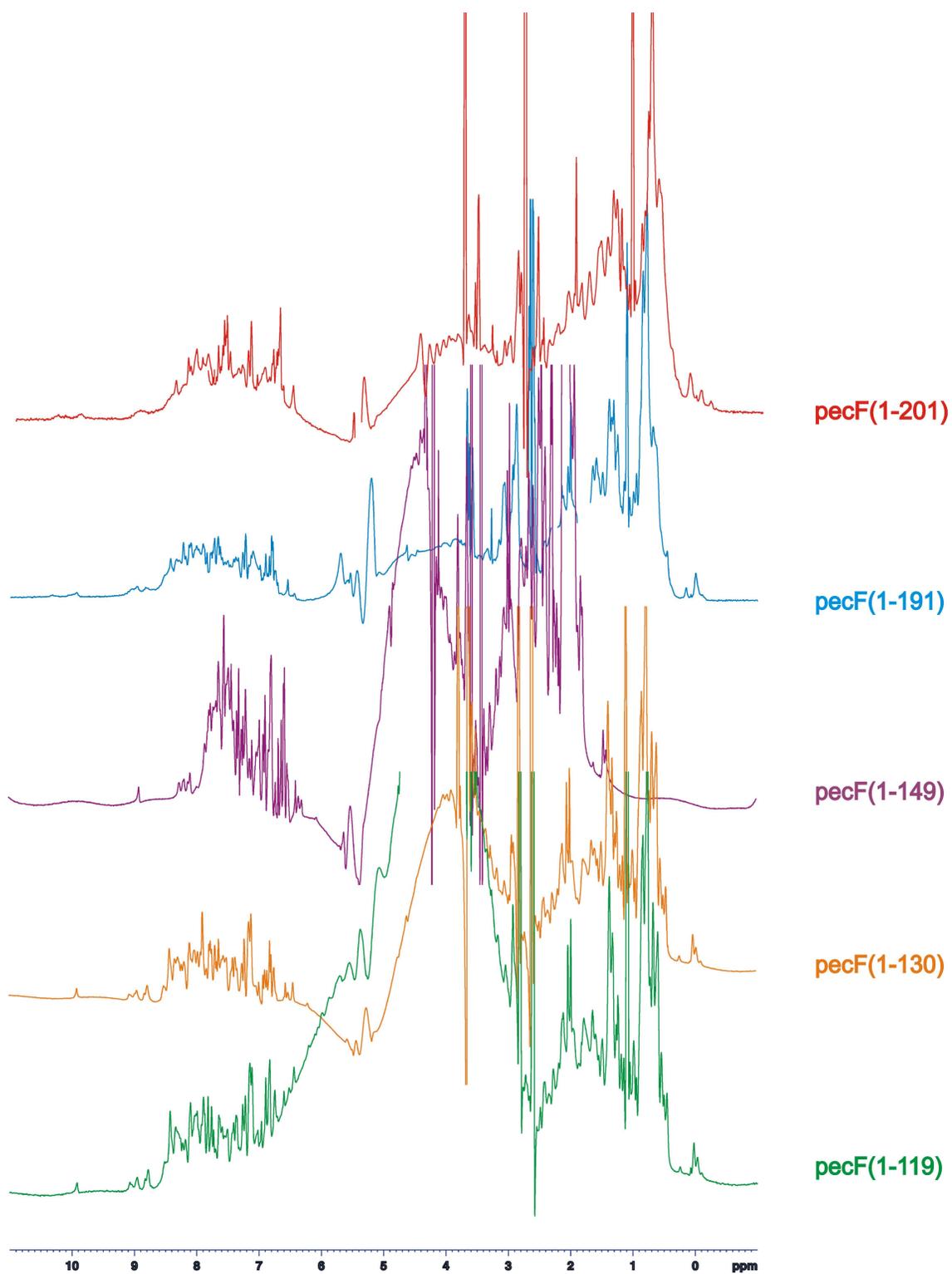
Figure 3.2.2. 1D-NMR proton spectra of various constructs of PecF.

3.2.1.3. Limited proteolysis

In order to identify the stable core of the PecF protein, a degradation study was performed. The protein incubated at room temperature for 72 h yielded a 25 kDa-long fragment, which was subjected to gel filtration in order to eliminate heterogeneity, and screened for crystallization.

PecF seemed to be resistant for the short-termed incubation at 4°C, however, a prolonged (3 months) incubation yielded a stable, completely folded fragment (Figure 3.2.3B), which was identified as covering residues 6-109. This fragment was cloned as an independent construct (PecF(6-109)) and a 1D proton NMR spectrum of the new construct was equivalent to the product of the of prolonged degradation (Figure 3.2.3C). PecF(6-109) was screened for crystallization and studied for its functional properties.

3.2.1.4. Functional properties of PecF(6-109)

Reconstitution studies to determine catalytic activity of PecF(6-109) were performed following the previously established protocol (Böhm et al., 2007). Reconstitution reaction requires the apoprotein phycoerythrocyanin A (PecA), enzyme phycoerythrocyanin lyase subunits E and F (PecE and PecF), and the chromophore phycocyanobilin (PCB) in 0.5 M Tris pH 6.0. The previously established optimal conditions include also presence of 3 mM $MnCl_2$, 5 mM β-mercaptoethanol and 10% Triton X-100. The reconstitution reaction was performed at 37°C, in dark, with gentle agitation, with the sample removed from the mixture every 90 min and examined spectroscopically. According to Böhm et al. (2007), the formation of the mature form of α-PEC can be detected by appearance of the absorption peak at 570-

575 nm. No additional peak was observed when the full length PecF was substituted
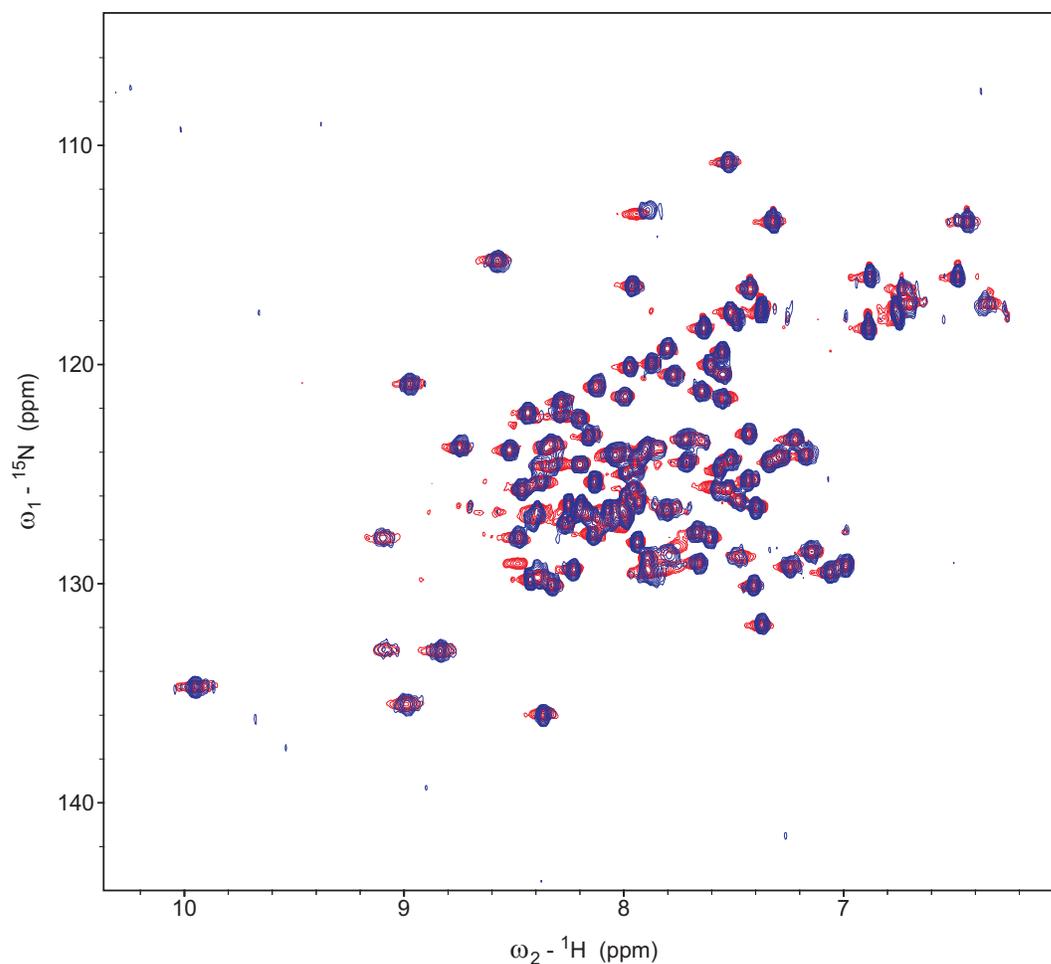
with PecF(6-109).
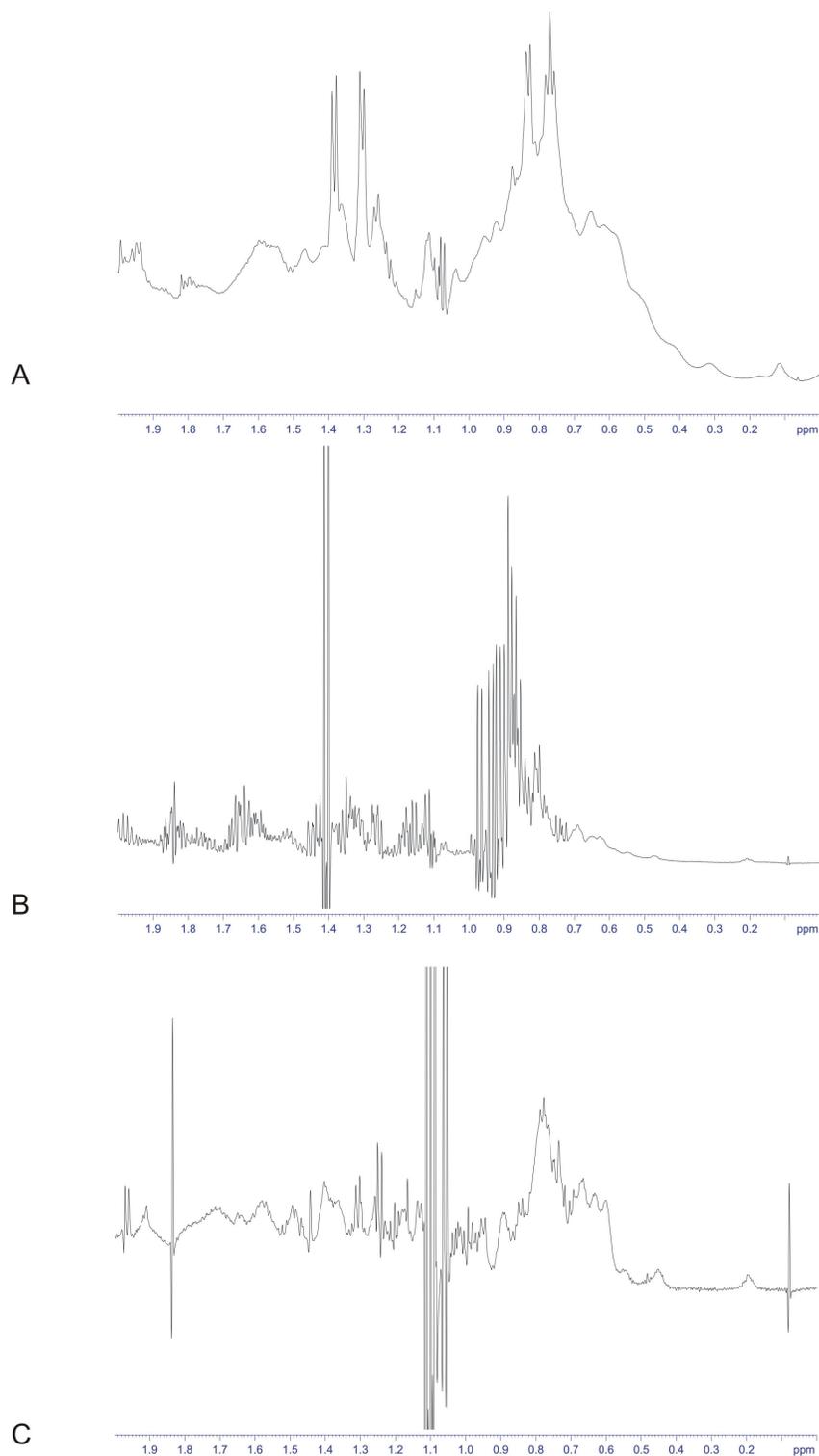
Figure 3.2.3. 1D-NMR spectra of PecF (**A**), the degraded PecF (**B**), and the PecF core (**C**).

To determine the cause of lack of the catalytic activity, binding studies with PecE were performed. Mobility shift assays were performed after PecF(6-109) was mixed with PecE in approximately equimolar concentrations and subjected to gel filtration on Sephadex 75. Since the proteins did not coelute, it can be reasoned that they did not interact with each other (Figure 3.2.4).
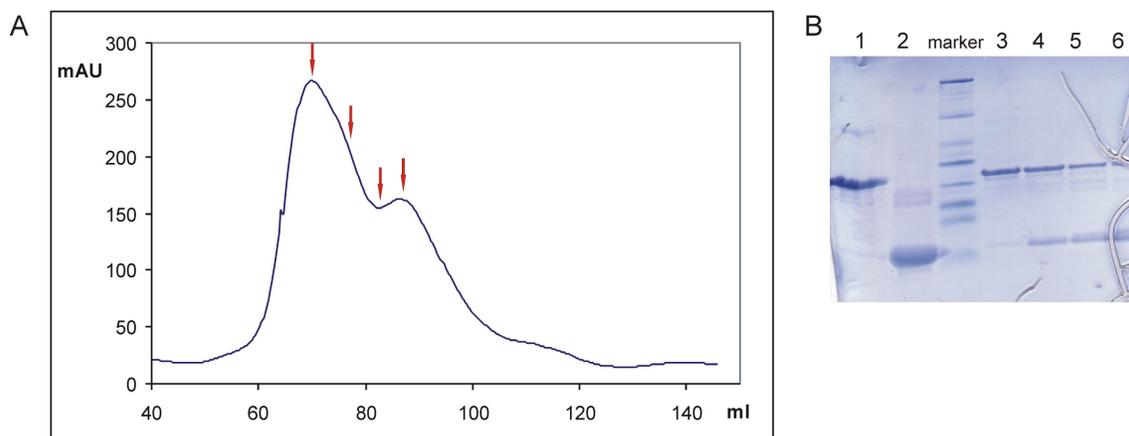
Figure 3.2.4. The gel filtration binding assay. (**A**) Elution profile, (**B**) SDS-PAGE image of the proteins prior to gel filtration (1 – PecE, 2 – PecF(6-109)) and selected fractions (3-6).

In order to exclude weaker interaction, a more sensitive method, namely HSQC NMR, was employed. The ^{15}N-uniformly labeled PecF(6-109) was produced and 2D HSQC NMR spectrum was recorded. After addition of the unlabelled PecE, no difference in the spectrum could be detected indicating lack of the interaction (Figure 3.2.5).

Figure 3.2.5. Superimposed 2D HSQC NMR spectra of PecF(6-109) alone (red) and the mixture of PecF(6-109) together with PecE (blue).

3.2.2. PecE

3.2.2.1 Construct design and crystallization

The phycoerythorocyanin lyase subunit E (PecE) isolated form cyano-bacterium *Mastigocladus laminosus* and cloned into the pET30a vector was a gift from Dr. S. Böhm. The PecE protein was expressed, purified and screened for crystallization as described for PecF. Table 3.2.3 summarizes constructs and the screening.

Table 3.2.3

Construct	Vector	Primers	MW (kDa)	Screening	Cleavage of tag	
					Uncleaved	Cleaved
1-269	pET30a	-	35.3 (30.3)	+	+	+
27-218	pET46	12,13	23.8	- (insoluble)		
27-218_Xa	pET30a	14,15,16,17	36.2 (21.6)	- (unstable)		
1-226	pET46	18,19	26.8	+	+	-
1-167	pET46	18,20	20.0	+	+	-

Molecular weight is calculated for the protein containing a tag. Values in brackets represent molecular weights after removal of the tag or fragments of the protein (for PecE(27-218)_Xa).

Based on the SMART (Simple Modular Architecture Research Tool, http://smart.embl-heidelberg.de) analysis, a construct comprising residues 27-228, excluding the region recognized as unstructured, was designed and cloned into pET46. Under standard, as well as modified (addition of glycerol or ethanol to the growth medium) conditions, the protein could be obtained in the insoluble fraction only. To overcome this difficulty, factor Xa cleavage sites were introduced, after residues 26 and 228. This approach enables the production of soluble proteins that would afterwards be stripped off of the unwanted fragments by proteases. A soluble protein could be obtained; however it proved to be unstable upon cleavage with factor Xa.

3.2.2.2. Limited proteolysis

Since degradation of PecE could not yield any well-defined product, the limited proteolysis approach was employed. Two distinguished bands detected after incubation of PecE with trypsin were identified by mass spectroscopy as covering residues 1-167 and 1-226 (Figure 3.2.6).
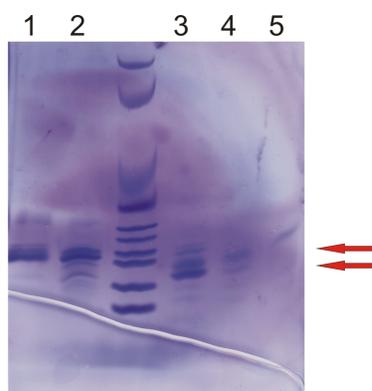
Figure 3.2.6. SDS-PAGE image of samples from trypsin digestion of PecE at 37ºC. 1 – PecE, 2 – Pec + trypsin, 3 - 5 – samples after 5, 30 and 60 min of digestion, respectively. Arrows indicate positions of fragments.

Constructs comprising these residues were cloned into the pET46 vector, expressed, and screened for crystallization. 1D-NMR spectra of the screened constructs are shown in Figure 3.2.7. It can be noticed that overall folding of the protein improves along with the shortening of the construct.
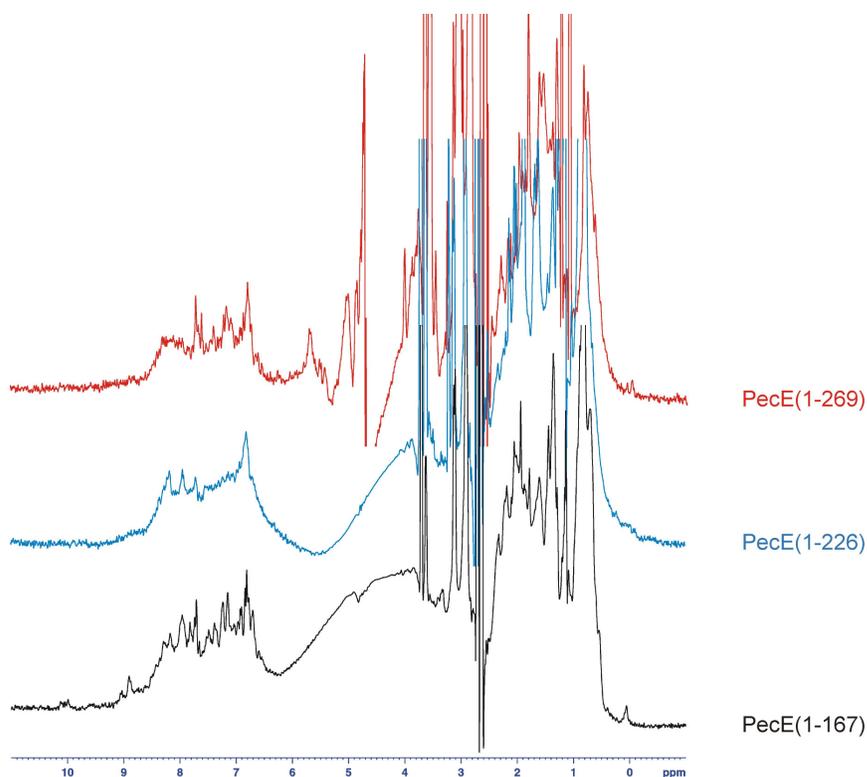

Figure 3.2.7. 1D-NMR proton spectra of various constructs of PecE.

3.2.3. Protein complexes of the phycoerythrocyanin lyase system

3.2.3.1 PecE, PecF, and PecA complexes

PecE and PecF being subunits of the phycoerythrocyanin lyase, are known to interact with each other. Purified proteins were mixed and subject to gel filtration in order to separate the formed complex from unbound subunits. The same procedure was applied for the trimeric complex, additionally including α-subunit of phycoerythrocyanin (α-PEC) (Figure 3.2.8). Both complexes were screened for crystallization, as described above.
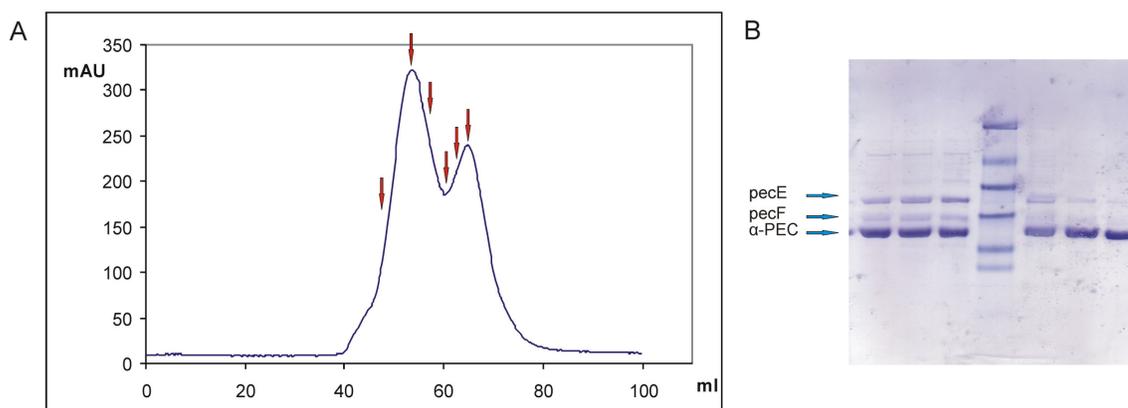
Figure 3.2.8. (A) Gel filtration of the PecE – PecF – α-PEC complex (red arrows indicate fractions chosen for electrophoresis), (**B**) SDS-PAGE image of chosen fractions. Blue arrows indicate positions of proteins´ bands in the gel.

3.2.3.2 The Z form of α–PEC

α–PEC can occur in two isomerization forms: E, for which the 3D structure is available (Schimdt et al., 2005), and in the Z-form. To obtain crystals of the Z-form, α–PEC were reconstituted as described above (using the full-length PecF instead of its truncated version) and left until the spontaneous degradation stripped off the first N-terminal helix. This form, called mini α–PEC, was purified by gel filtration and

subjected to crystallization screening. Plates were irradiated with the 500 nM light to induce complete photoconversion to the Z-form.

Additionally, according to previous observations, the already established conditions were optimized namely: 0.2 M magnesium sulphate, 0.1 mM a potassium phosphate, 25% isopropanol; and 0.2 M magnesium sulphate, 0.1 M potassium phosphate, 25% isopropanol, 25% PEG 4000. Crystals could be obtained, but due to large quantities of volatile isopropanol in the drops, they were difficult in handling and were unstable (Figure 3.2.10B).

3.2.3.3 P641 adduct

PecA adds autocatalytically PCB, yielding a low absorbance, low fluorescence PCB : PecA adduct, which can be transformed through the action of PecE into a high absorbance, high fluorescence PCB : PecA adduct, termed P641 (Böhm et al., 2007).

The reconstitution of the P641 adduct was carried out according to the protocol of Böhm et al. (2007). Briefly, the PecE and PecA proteins were mixed in equimolar ratio in 0.5 M Tris, pH 6.0, with addition of the chromophore (PCB), 3 mM Mn^{2+}, 5 mM β-mercaptoethanol, and incubated for 3 h at 37ºC, in the dark. Afterwards, the sample was purified on the NiNTA resin (to remove the unbound chromophore) and subjected to gel filtration (Figure 3.2.9). Two subfractions could be obtained: one containing P641 alone and the second, P641 in the complex with PecE.
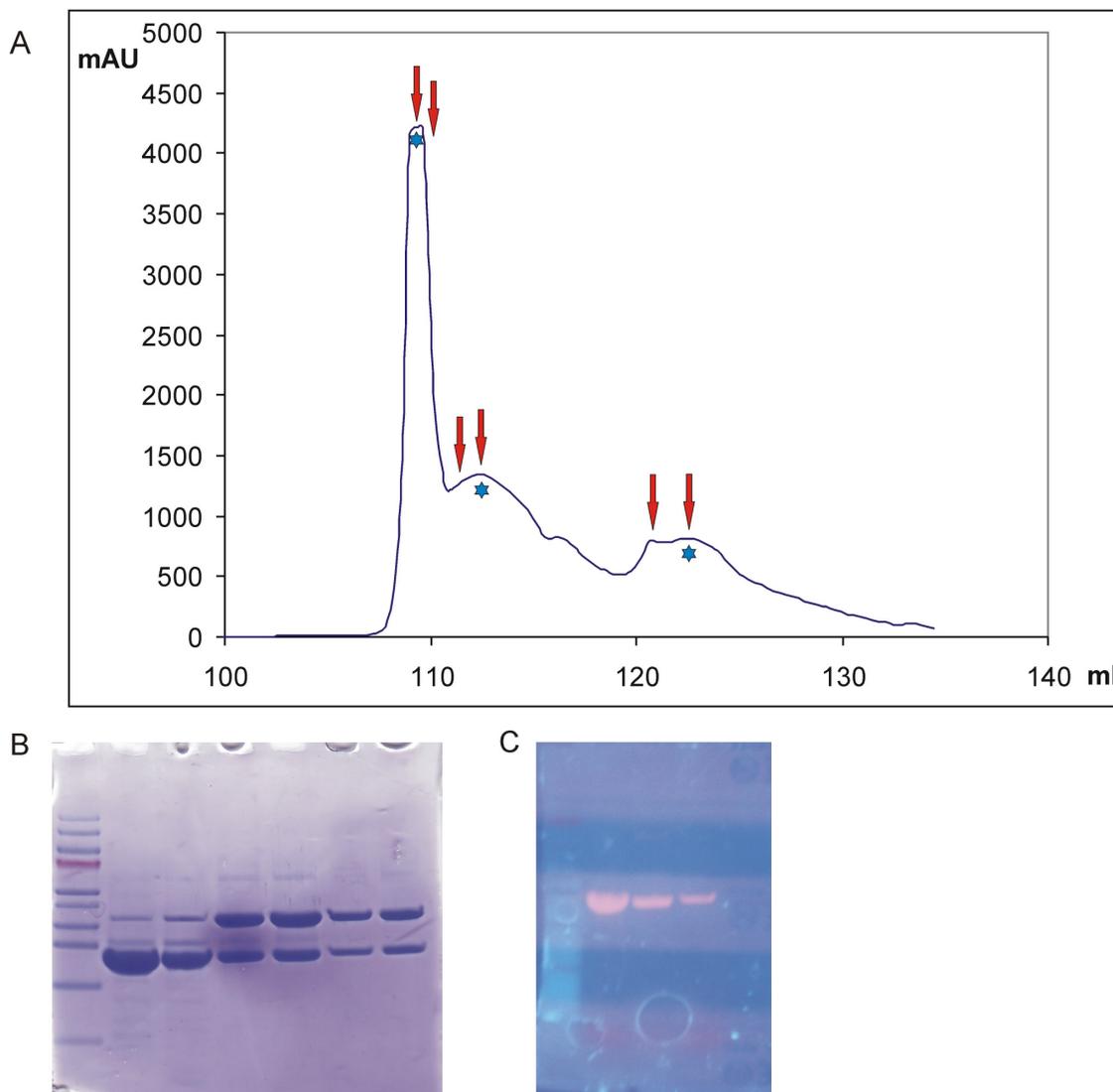
Figure 3.2.9. (**A**) Gel filtration of the P641 adduct (red arrows indicate fractions chosen for SDS-PAGE stained with Coomasie Blue; blue asterixes indicate fractions chosen for SDS-PAGE stained with Zn^{2+}), (**B**) SDS-PAGE image of fractions stained with Coomasie Blue, (**C**) SDS-PAGE image of fractions stained with Zn^{2+} (fluorescence of the chromophore is visualized).

Due to lack of the isomerizing subunit (PecF) in the reconstitution mixture, the apoprotein contains phycoerythrobilin instead of phycoviolobilin. This adduct can be considered as an intermediate and therefore it remains bound to PecE throughout the NiNTA and gel filtrations. Adduct P641, as well as the complex of the adduct and PecE, were screened for crystallization. The reconstitution procedure was repeated

for proteins deprived of their respective His-tags and the resulting samples were subjected to screening as well. An uncleaved adduct P641 produced cluster-like crystals that could not be optimized for better quality (Figure 3.2.10A).
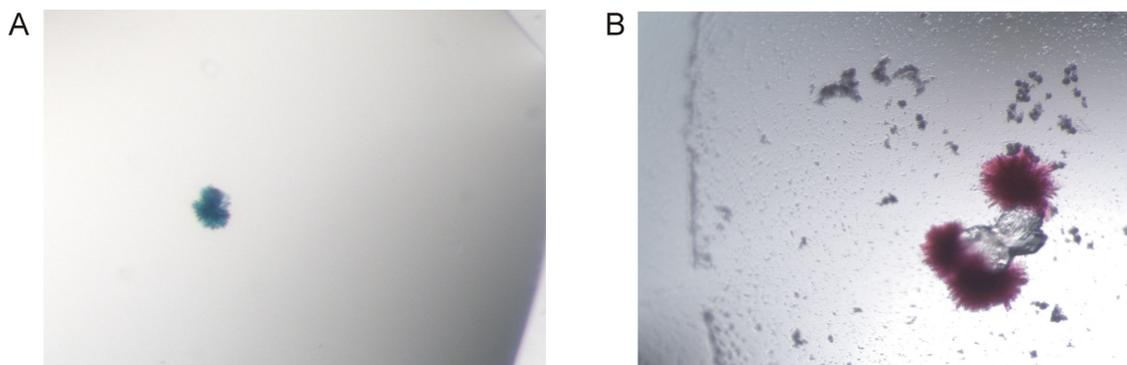
Figure 3.2.10. Images of crystal-like structures of the P641 adduct (**A**) and mini α-PEC (**B**).

3.2.4. Cpc lyases E/F

The *Synechococcus* PCC7002 and *Anabaena* PCC7120 strains were ordered from the Institute Pasteur, France. Genes of the CpcE and CpcF lyases were fished out using the cyanbacteria that were lyzed by repeated freeze-thaw cycles using liquid nitrogen. CpcE and CpcF from PCC7002, along with CpcF from PCC7120, could be isolated and cloned into the pET46 vector.

Proteins were expressed, purified and screened for crystallization as described before, and, in the case of CpcE/F from PCC7002 also in an equimolar complex. All proteins gave crystals in multiple conditions (Table 3.2.4) and are shown in Figure 3.2.11.

Since no structures of a similar fold could be detected in the database search (Protein Homology/Analogy Recognition Engine - http://www.sbg.bio.ic.ac.uk/phyre) the phase problem had to be solved through the selenomethionine-labeling approach. The selenomethionine labeled protein was obtained following a standard

procedure with the exception of the BL21(DE3) cells substituted by the B834(DE3) cells and the LB medium being substituted by the defined modified M9 medium with the methionine replaced by selenomethionine.

Table 3.2.4. Crystallization conditions

Condition	Protein	Composition	Temp	Diffraction
Anions 33	CpcF7120	0.1 M MES pH 6.5, 3.5 M sodium formate	4ºC	
Anions 34	CpcF7120	0.1 M MES pH 6.5, 1.75 M sodium formate	4ºC	* (native)
Anions 38	CpcF7120	0.1 M HEPES pH 7.5, 1.75 M sodium formate	4ºC	
Anions 44	CpcF7120	0.1 M Tris pH 8.5, 1.75 M sodium formate	4ºC	* (native)
Anions 55	CpcF7120	0.1 M MES pH 6.5, 3.5 M sodium nitrate	4ºC	7.0 (SeMet)
Anions 56	CpcF7120	0.1 M MES pH 6.5, 1.75 M sodium nitrate	4ºC	
Anions 62	CpcF7120	0.1 M HEPES pH 7.5, 1.75 M sodium nitrate	4ºC	
Anions 67	CpcF7120	0.1 M Tris pH 8.5, 3.5 M sodium nitrate	4ºC	7.0 (SeMet)
pHClear 10	CpcF7120	0.1 M HEPES pH 7.0, 2 M sodium chloride	4ºC	* (native)
pHClear 11	CpcF7120	0.1 M Tris pH 8.0, 2 M sodium chloride	4ºC	
pHClear 12	CpcF7120	0.1 M Bicine pH 9.0, 2 M sodium chloride	4ºC	4.5 (SeMet)
pHClearII 12	CpcF7120	0.1 M Bicine pH 9.0, 1 M lithium chloride, 10% PEG 6000	4ºC	* (SeMet)
Index 9	CpcF7120	0.1 M BIS-Tris pH 5.5, 3 M sodium chloride	4ºC	7.0 (SeMet)
Classics 12	CpcF7120	10% ethanol, 1.5 M sodium chloride	4ºC	5.0 (SeMet)
Wizard 92	CpcF7120	0.1 M acetate pH 4.5, 0.2 M lithium chloride, 2.5 M sodium chloride	20ºC	1.9 (native)
Classics 50	CpcF 7002	0.1 M MES pH 6.5, 10% dioxane, 1.6 M ammonium sulfate	4ºC	2.2 (native)
Classics 68	CpcF 7002	0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5, 20% PEG 8000	4ºC	
Classics 95	CpcF 7002	0.1 M HEPES pH7.5, 8% ethylene glycol, 20% PEG 10000	4ºC	
Classics 88	CpcE 7002	0.2 M magnesium chloride, 0.1 M Tris pH 8.5, 30% PEG 4000	20ºC	
Index 84	CpcE 7002	0.2 M magnesium chloride, 0.1 M HEPES pH 7.5, 30% PEG 3350	20ºC	2.8 (native) 3.5 (SeMet)
Index 85	CpcE 7002	0.2 M magnesium chloride, 0.1 M Tris pH 8.5, 30% PEG 3350	4ºC, 20ºC	3.5 (SeMet)

SeMet – a selenomethionine-labeled crystal, * - non diffracting

Several crystals of the selenomethionine labeled CpcF PCC7120 and CpcE PCC7002 were grown, harvested and measured on the PXII beamline of the Swiss Light Source (Villigen, Switzerland). None of the SeMet labeled crystals tested so far diffracted well enough to enable the phasing and structure solution. A native CpcF7120 from the Wizard 92 condition and the native CpcF7002 from the Classics 50 condition diffracted to 1.9 and 2.2 Å respectively, but none of these could be repeated with the selenomethionine labeled protein.
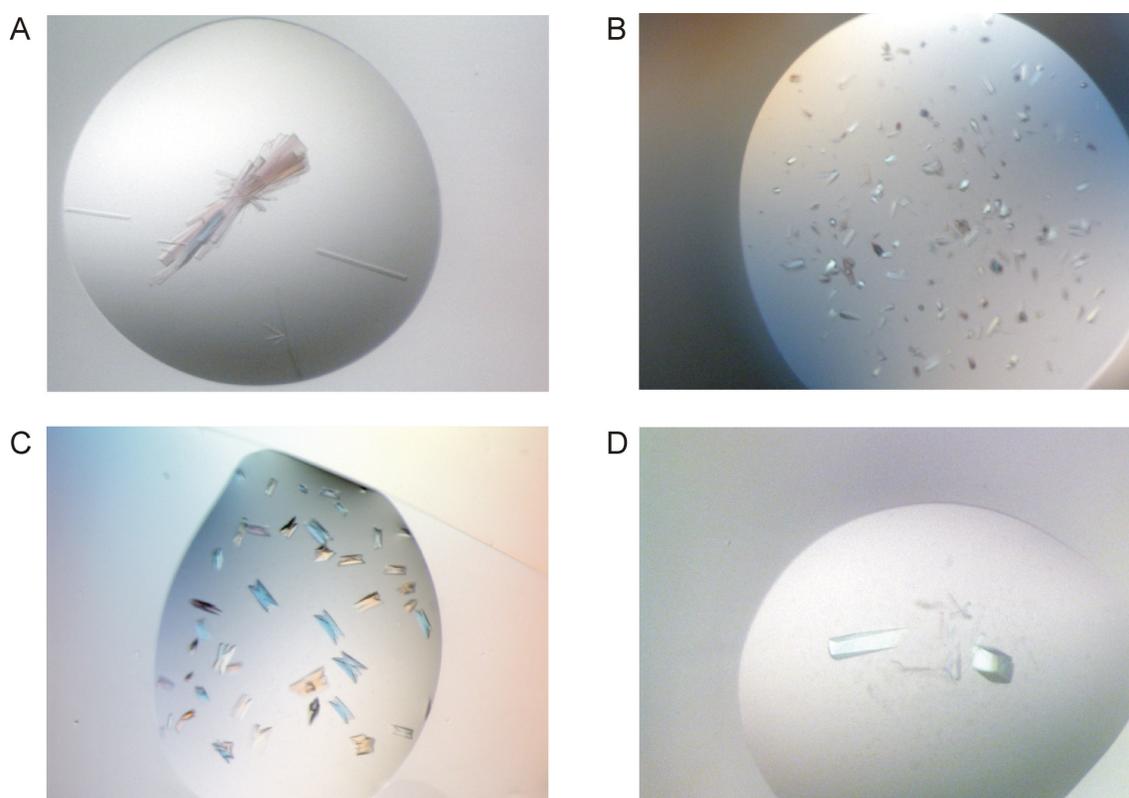
Figure 3.2.11. Crystal images of CpcF 7120 native in pHClear 12 (**A**) and selenomethionine-labeled in Anions 67 (**B**), CpcE 7002 in Index 84 (**C**) and CpcF 7002 (**D**).

3.3 Site-specific integrases

3.3.1. HPI integrase

3.3.1.1 Construct design, cloning and purification of the HPI integrase

The N-terminal domain of the HPI integrase from *Yersinia pestis* was cloned into the pET21b vector. The initial construct, named HPI_96, comprised residues 1-96, with borders of the construct established based on the secondary structure prediction. Research on the phage lambda already showed that the arm-type binding domain is separated from the C-terminal, catalytic domain by a short linker. Our construct comprised thus the N-terminal domain, the whole linker and the residues covering the following α-helix (3.3.1A).

A   HPI_96

```
1                     21                    41
MSLTDAKIRTLKPSDKPFKVSDSHGLYLLVKPGGSRHWYLKYRISGKESRIALGAYPAIS
----HHHHHH------EEEEE----EEEEEE----EEEEEEEEE--EEEEE--------
61                    81
LSDARQQREGIRKMLALNINPVQQRAAERGSRTPEK
HHHHHHHHHHHHHHH----HHHHHHHHHHHHH--
```

B   HAI7_100

```
1                     21                    41
MSLTDSKVKNAKSLEKEYKLTDGFGMHLLVHPNGSKYWRLSYRFEKKQRLLALGVYPAVS
----HHHHHH------EEEEE----EEEEEE----EEEEEEEE----EEEEE--------
61                    81
LADARQRRDEAKKLLAAGIDPSAKKQADNKTIQEKRNNTR
HHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHH---
```

C   HAI13_100

```
1                     21                    41
MSLTDTKVKNAKPSEKVVKLTDGFGLYLLVHTNGSKYWQLGYRFEGKQKVFSIGVYPAVS
----HHHHHH------EEEEE----EEEEEE----EEEEEEEE----EEEEE--------
61                    81
LANARQRRDEAKKLLAAGVDPSAKKRADNKSVQEKRNNTR
HHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHH---
```

Figure 3.3.1. The secondary structure prediction for amino acids 1-96 of the HPI integrase (**A**), 1-100 of the HAI7 integrase (**B**) and 1-100 of the HAI13 integrase (**C**). Linker residues are marked in green, catalytic domain residues are marked in blue. H, indicates the predicted α-helix, while E stands for a β-strand.

The construct was expressed in the *E. coli* strain BL21(DE3) in the LB medium, with induction at 27°C overnight. Afterwards cells were spinned down, re-suspended in the lysis buffer and disrupted by sonication. Clarified solution was applied onto the NiNTA resin, and after washing with the wash buffer, eluted by high concentration of imidazole. Eluted fractions were subjected to the gel filtration in a buffer suitable for crystallization.

As examined by SDS-PAGE electrophoresis, the HPI_96 protein turned out to be unstable, since a smaller band (of an approximate mass of 10 kDa) appeared spontaneously after 2 days of storage. Based on mass the spectrometry data, this band was identified as containing two fragments spanning residues 1-76 and 1-83. That indicated that the α-helix belonging to the catalytic domain was cleaved with the cleavage occurring in the region of the linker.

New construct covering residues 1-80 (named HPI_80), i.e. the N-terminal domain and the linker exclusively, was designed and cloned into the pET21b vector. The protein, which was expressed and purified as previously, was eluted from the gel filtration in form of dimer as well as monomer (Figure 3.3.2). HPI_80 proved to be folded by 1D-NMR.
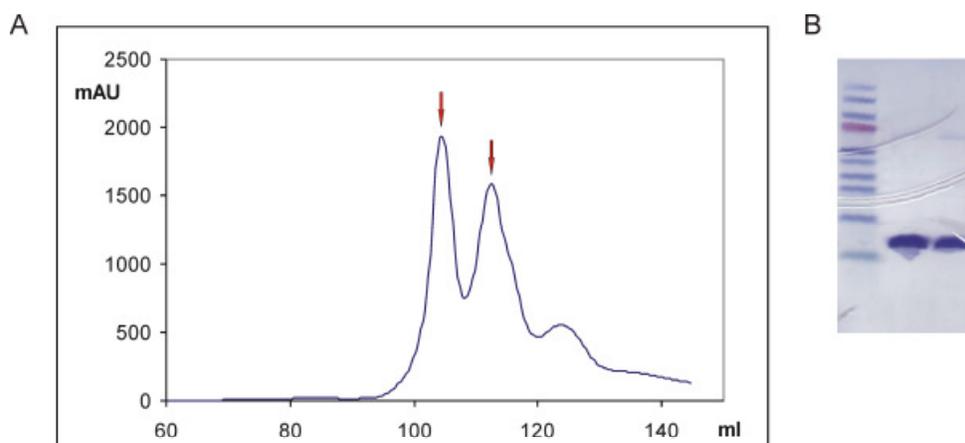
Figure 3.3.2. (**A**) Gel filtration of HPI_80. Red arrows indicate fractions selected for electrophoresis, (**B**) SDS-PAGE image of chosen fractions.

3.3.1.2 Crystallization and structure determination

HPI_80 was screened for crystallization using 96-well sitting drop plates with commercially available factorials (Qiagen's Classics, PEGs MPDs, Cations, Anions, pHClear 1 and 2, and Hampton Research Index) at 20ºC and 4ºC. The screening gave rise to crystals in several conditions (Figure 3.3.2 and Table 3.3.1). Prior to the data collection all crystals were plunge-frozen in the mother liquor containing additionally 30% of MPD as a cryoprotectant.

Table 3.3.1. Crystallization conditions

Condition	Protein	Composition	Temp
Classics 7	HPI_80	0.1 M tri-sodium citrate pH 5.6, 20% isopropanol, 20% PEG 4000	4ºC
Classics 9	HPI_80	0.2 M tri-sodium citrate, 0.1 M sodium cacodylate pH 6.5, 30% isopropanol	4ºC
Classics 21	HPI_80	0.2 M ammonium sulphate, 0.1 M Tris pH 8.5, 30% MPD	4ºC
Classics 24	HPI_80	0.1 M tri-sodium citrate at pH 5.6, 35% (v/v) tert-butanol	4ºC, 20ºC
MPD 39	HPI_80	0.2 M sodium phosphate, 40% MPD	4ºC
MPD 43	HPI_80	0.2 M ammonium phosphate, 40% MPD	4ºC
pHClear 69	HAI7_100	3.2 M ammonium sulfate, 0.1 M MES pH 6.0	4ºC
Classics 82	HPI_80 / attP_21bp	0.01 M nickel chloride, 0.1 M Tris pH 8.5, 20% PEG 2000 MME	20ºC
Classics 95	HPI_80 / attP_21bp	0.1 M HEPES pH 7.5, 20% PEG 10000, 0.2 8% Ethylene glycol	20ºC
pH Clear 1	HPI_80 / attP_21bp	0.1 M citric acid, 65% MPD	20ºC
MPD 68	HPI_80 / attP_21bp HPI_80 / attP_15bp	0.1 M sodium acetate pH 6.5, 65% MPD	20ºC
MPD 75	HPI_80 / attP_15bp	0.02 M calcium chloride, 0.1 M sodium acetate pH 4.6, 15% MPD	20ºC

The crystals originated from the Classics 7 condition allowed for the collection of a dataset of the 1.9 Å resolution at a rotating anode laboratory source. The collected data were integrated, scaled and merged by XDS and XSCALE programs (Kabsch, 1993). Crystals belonged to the space group $P4_12_12$ and contained one molecule per an asymmetric unit.
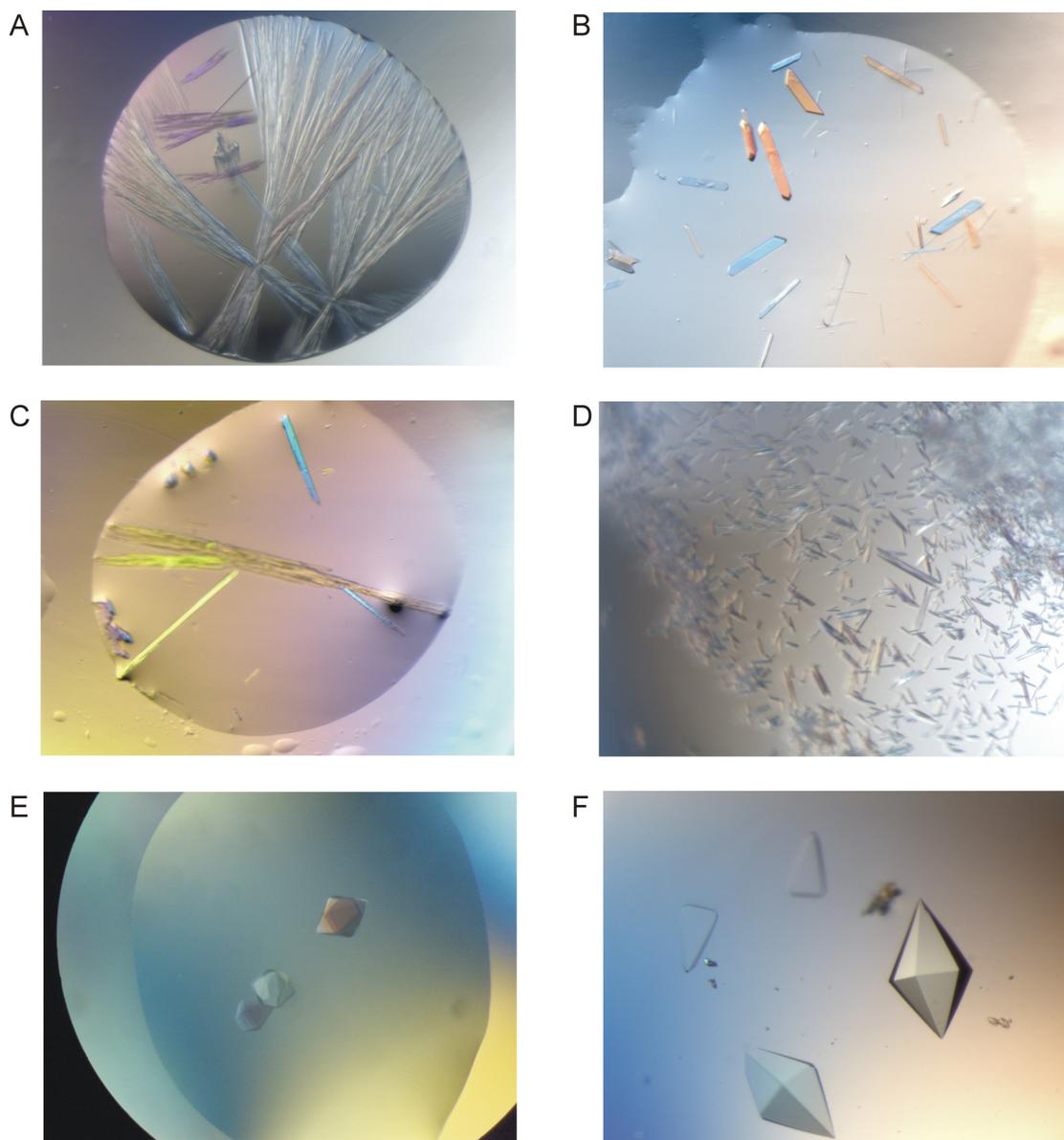
Figure 3.3.2. Crystal obtained in the commercially available factorials: (**A**) HPI_80 in Classics 7, (**B**) HPI_80 in Classics 9, **C.** HPI_80 in Classics 21, (**D**) HPI_80 in Classics 24, (**E, F**) HAI7_100 in pHClear 1 69.

Several structures of proteins binding DNA by means of the β-sheet are known, namely: the N-terminal domain of the transposon 916 integrase, the arm-type binding domain of the phage lambda integrase, and the GCC-box binding domain of the ethylene-responsive element binding protein from *Arabidopsos thaliana* (Allen et al., 1998; Wojciak et al., 1999a, Wojciak et al., 1999b). Due to the expected structural similarity with the HPI integrase, the respective structures (PBD entries: 1B69, 1KJK, 1GCC) were used as probes for the molecular replacement by the Molrep program of the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Unexpectedly, none of these structures allowed the building of the model of the N-terminal domain of HPI. Since all abovementioned structures were NMR solution structures, they contained vast regions of high mobility that could interfere with the molecular replacement procedure. Utilization of models deprived of flexible regions did not provide any improvement.

Therefore, in order to solve the phase problem, the multiwavelength anomalous diffraction (MAD) approach was attempted. The selenomethionine labeled protein was obtained following standard procedures with the exception of the BL21(DE3) cells substituted by the B834(DE3) cells and the LB medium substituted by the modified M9 medium with the methionine replaced by selenomethionine. The selenomethionine-labeled crystals were harvested from the optimized Classics 9 condition with the cryoprotection by MPD.

The diffraction data were measured on the PXII beamline of Swiss Light Source (Villigen, Switzerland). Data were indexed, integrated and scaled with the XDS package (Kabsch, 1993). The selenomethionine derivative was measured at 0.9796 Å for the peak, 0.9796 Å for the inflection point, 0.972 Å for the high remote and 0.9875 Å for the low remote datasets; all datasets were collected up to 2.0 Å.

Two selenomethionine crystals were measured and corresponding datasets merged to obtain higher redundancy. The datasets were high quality and showed strong anomalous signals. Anomalous scaterers were found using the SHELXD software (Schneider et al., 2002). Interestingly, two sites, instead of the expected one site, were found. After map analysis, it turned out that the Se atom is very close to a "special position" and the selenomethionine sidechain assumes different conformations in each asymmetric unit, thus locally breaking the crystallographic symmetry (this is also the case in the native data for the sulphur atom). Two initial atom positions were refined using the autoSHARP software package (de La Fortelle and Bricogone, 1997). The resulting phases were improved by the DM program (Bailey, 1994) and used for an automated model building with the Arp/Warp software (Perrakis et al., 1999). The resulting model of about 80% completeness, was inspected and finished manually with the Xfit program (McRee, 1999). Restrained refinement enforced by the Refmac5 software was then performed using a newly collected native data (diffraction to 1.3 Å) and followed by addition of water molecules by Arp/Warp (Lamzin and Wilson, 1993). Final electron density maps were of high quality; there were however no interpretable densities for solvent exposed sidechains of Lys12 and Arg43. These side chains were removed from the model. The final R-factor was 20.8% and Rfree 23.5%. Data collection, phasing, and refinement statistics are summarized in Table 3.3.2.
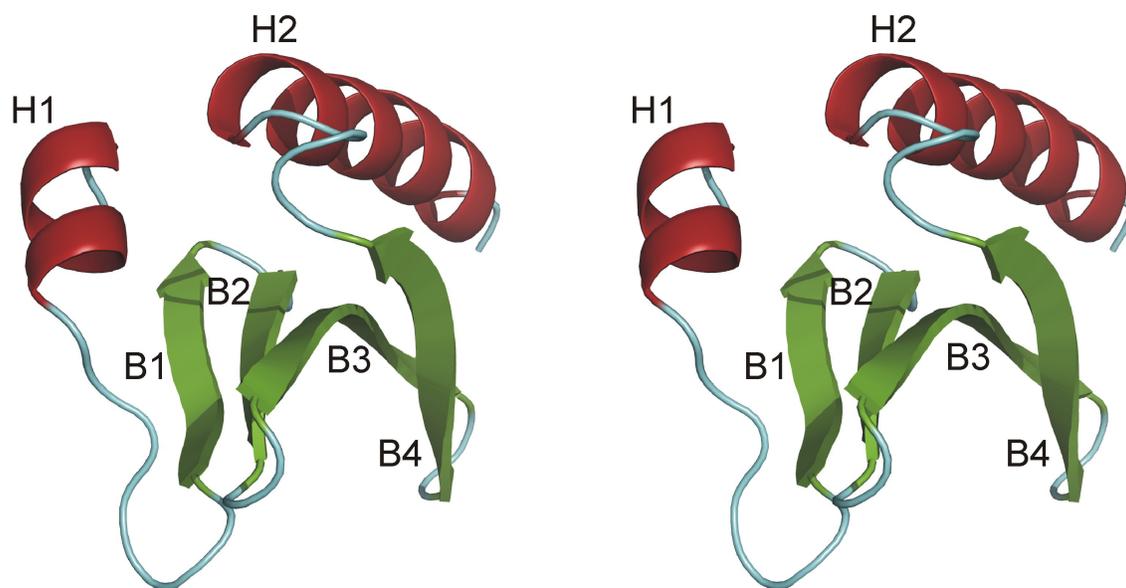
3.3.1.3 Structure of the HPI integrase

3.3.1.3.1 Overall structure

The solved structure comprises a four-stranded antiparalell β-sheet preceded and followed by α-helices. The N-terminal α-helix H1 covers residues Asp5-Thr10

and is positioned perpendicularly in respect to the C-terminal, larger helix. H1 is followed by a 7 residue-long loop connecting it to the first β-strand B1 (Phe18-Ser23). β-Strands B1, B2 (Leu26-Lys31), B3 (Gly34-Ile44), and B4 (Lys47-Ala55) are connected by short turns comprising 2 amino acids, all of them belonging to the I' category according to Sibanda et al. (1989).

β-strands are parallel in following pairs: B1 with B2 and B3 with B4. First two strands (B1 and B2) form an angle of about 45º in respect to strands B3 and B4. The central part of the β-sheet connects to the second α-helix H2 (Leu61-Ala76) by a 5 aminoacid-long loop. This helix is positioned almost parallel to strands B3 and B4. The β-sheet together, with helix H2, forms a L-shaped hydrophobic core of the protein.

Figure 3.3.3. Crystal structure of the arm-type binding domain of the HPI integrase in ribbon representation.

Table 3.3.2. Data collection and refinement statistics for the HPI integrase

	Native	peak	inflection	high remote	low remote
Space group			$P4_12_12$		
Cell constants (Å)					
a			48.75		
b			78.75		
c			74.20		
	Native		*SeMet*		
Dataset		*peak*	*inflection*	*high remote*	*low remote*
Resolution range (Å)	20-1.3	50-2.0			
Wavelength (Å)	0.9873	0.9796	0.9799	0.9720	0.9875
Observed reflections	391656	196899	196596	196716	196899
Unique reflections	19463	11516	11505	11510	11516
Whole resolution range:					
Completeness (%)	85	98.4	98.3	98.4	98.4
R_{merge}	3.6	1.7	1.7	2.6	2.8
$I/\sigma(I)$	29.07	45.32	43.69	46.44	45.32
Last resolution shell:					
Resolution range (Å)	1.3-1.4	2.0-2.1			
Completeness (%)	65.5	96.6	96.3	96.8	96.6
R_{merge}	10.3	2.9	2.9	2.6	2.8
$I/\sigma(I)$	8.98	29.85	28.42	31.62	29.85
Phasing					
Number of sites found/present	-	2/2 (each Se site as 50% occupancy)			
Phasing power anomalous/isomorphous	-	4.2/1.622			
FOM	-	0.65			
Refinement					
No. of reflections			18404		
Resolution (Å)			10 – 1.3		
R-factor (%)			20.8		
R_{free} (%)			23.5		
Average B ($Å^2$)			16.7		
R.m.s bond length(Å)			0.008		
R.m.s. angles (°)			1.28		

Content of asymmetric unit		
No. of protein molecules		1
No. of protein residues/atoms		77/612
No. of solvent atoms		125
No. of heavy atom sites identified	-	2
Ramachandran statistics		
Most favored regions (No./%)		59/91
Additionally allowed regions (No./%)		6/9
Generously allowed regions (No./%)		0/0
Disallowed regions (No./%)		0/0

3.3.1.3.2 Comparison with the structures of the three-stranded β-sheet DNA binding proteins

All known proteins that bind their DNA recognition sites by means of β-sheets contain three β-strands. The HPI integrase differs from those previously published by a prominent, extra β-strand, which is an integral part of the β-sheet (Figure 3.3.4). Strand B1 can be considered as an "additional" due to its orientation in respect to the rest of the sheet. Further, all β-strands in the structure of the arm-type binding domain of the HPI integrase are significantly longer (consisting of 6-11 amino acids) than in any other arm-type DNA binding domains (consisting of 3-6 amino acids). This translates into a number of residues projecting from the β-sheet in the direction of the potential DNA binding cavity with as many as 13 residues in HPI integrase (the Tn916 transposon integrase having 9 of them, while the lambda integrase only 7).
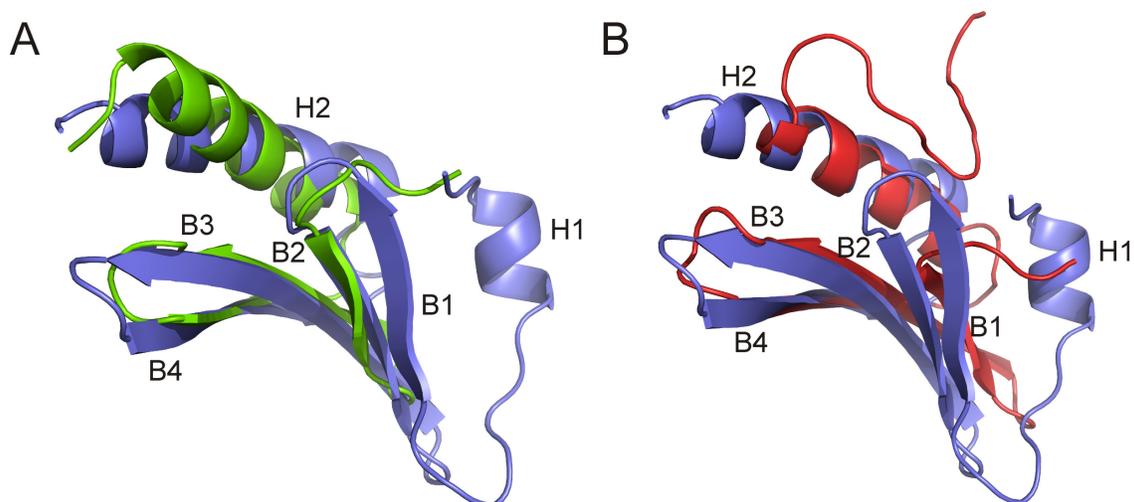
Figure 3.3.4. The structure of the HPI integrase (blue) best fitted to those of the lambda integrase (**A**) and the GCC-box binding protein (**B**).

The other distinguished feature of our integrase domain is its structured, helical N-terminus positioned perpendicular to the large, C-terminal helix. Among all known integrases, a small N-terminal helix is present only in a newly described structure of the lambda integrase in complex with DNA (Fadeev et al., 2009). The positioning of the helix is, however, different from that of the HPI integrase, with the N-terminus of lambda integrase pointing away from the main body of protein. Since the structure of the N-terminus of the lambda integrase proved to be significantly different in free versus DNA-bound states, at this stage it could not be excluded that the N-terminal helix of the HPI integrase changes the conformation upon binding to the attachment site.

In addition, at this stage it was still uncertain whether the mode of the DNA binding is conserved between the integrases possessing three-stranded β-sheets and the HPI integrase. The space created by a concave structure beneath the β-sheet is however large enough in the HPI integrase to fit the major groove of DNA. Furthermore the charge distribution in HPI integrase is similar to other arm-type

binding proteins with positive charges gathered on the concave surface of the β-sheet (Figure 3.3.16).

3.3.2. HAI integrases

3.3.2.1 Construct design, cloning and purification

The integrases from another group of asn tDNA-associated genomic islands, namely HAIs (Horizontally Acquired genomic Islands) from the plant pathogen *Erwinia carotovora* were examined for their biochemical activity (Antonenka et al., unpublished results). HAI integrases share a degree of homology with the HPI integrase. The HPI sequence, when compared with HAI7 and HAI13, shows ca. 48% identity, irrespectively whether a full length or the N-terminal domain (amino acids 1-80) is taken into account. HAI7 and HAI13 integrases, being much closer related, share as much as 90,7% of identity in overall sequence. This value drops, however, significantly when only first 80 aminoacids are compared (Table 3.3.3). Since integrases HAI 7 and 13 are known to recognize distinctive DNA sequences this suggests that the specificity for the attachment site binding is localized within the N-terminal domain.

Table 3.3.3

Full Length		
	HAI7	HAI13
HPI	47.2%	47.4%
HAI7	-	90.7%
N terminal domain		
HPI	48.8%	48.8%
HAI7	-	77,5%

The HPI and HAI integrases recognize distinguished DNA sequences, therefore a comparison of their DNA binding cavities could be instructive in elucidation of specificity determinants of the DNA binding. Based on the crystallizable construct of the HPI integrase, equivalent constructs for HAI7 and HAI13 (spanning residues 1 to 80, Figures 3.3.1B and C) were designed and cloned in the pET21b. These proteins were expressed and purified according to the previously described protocol. While HAI13_80 was obtained in large amounts and subjected to crystallization screening, the HAI7_80 was found mostly in inclusion bodies. Therefore, a longer construct of the HAI7 integrase, covering residues 1-100 (Figure 3.3.1B), was prepared, expressed and screened for crystallization.

3.3.2.2 Crystallization and structure determination

Crystals of HAI7_100 grew initially in 0.1 M MES, pH 6.0, 3.2 M ammonium sulphate, and crystals optimized for the data collection were obtained from a drop containing 0.1 M MES, pH 6.3, 2.8 M ammonium sulphate (Figures 3.3.2E and F). The high-resolution dataset, up to 1.6 Å, was collected on the PXII beamline of the Swiss Light Source (Villigen, Switzerland). The crystals belonged to the space group $P6_5$ and contained one molecule per an asymmetric unit. The collected data were integrated, scaled and merged by the XDS and XSCALE programs. The structures were determined by molecular replacement using the Molrep program from the CCP4 suite with the HPI_80 structure used as a probe. The model was afterwards refined by Refmac5, and improved and revised manually by using XtalViev/Xfit. The Arp/wArp program was used to add solvent atoms. The final R-factor of the structure HAI7 integrase is 14.7% and R-free 21.2%. Data collection and refinement statistics are summarized in Table 3.3.4.

111

**Table 3.3.4. Data collection and refinement statistics for the HAI7 and HPI – DNA complex**

	HAI7	HPI-DNA complex
Space group	$P6_5$	$P2_12_12_1$
Cell constants (Å)		
a	52.56	60.900
b	52.53	69.490
c	79.47	83.740
Resolution range (Å)	6-1.6	2.5-19.8
Wavelength (Å)	1.0015	0.9873
Observed reflections	183210	83481
Unique reflections	15089	19792
Whole resolution range:		
Completeness (%)	91.6	81.5
R_{merge}	2.7	6.6
$I/\sigma(I)$	30.67	17.09
Last resolution shell:		
Resolution range (Å)	1.6-1.7	2.5-2.6
Completeness (%)	76	53.7
R_{merge}	10.7	18.4
$I/\sigma(I)$	9.31	6.13
Refinement		
No. of reflections	14113	13185
Resolution (Å)	6-1.6	2.5-9.8
R-factor (%)	0.1476	0.219
R_{free} (%)	0.2121	0.278
Average B (Å^2)	23.3	23
R.m.s bond length(Å)	0.024	0.016
R.m.s. angles (°)	2.051	3.58
Content of asymmetric unit		
No. of protein molecules	1	2
No. of protein residues/atoms	95/866	150/2356
No. of solvent atoms	130	46

Ramachandran statistics		
Most favored regions (No./%)	81/95	136/90.7
Additionally allowed regions (No./%)	4/5	12/8
Generously allowed regions (No./%)	0/0	2/1.3
Disallowed regions (No./%)	0/0	0/0

Final electron density maps were of high quality; there were, however, no interpretable densities for solvent exposed sidechains of Lys7, Lys 16, Lys 46, Lys 47, Arg49, Lys90, Gln93, Lys95, and Arg96. These side chains were removed from the model. Two clusters of residues with unidentified density can be noticed: in proximity to the loop between strands B3 and B4 (Lys 46, Lys 47 and Arg49) and at the C-terminus of the protein (Lys90, Gln93, Lys95, and Arg96). It can be assumed that these region display higher mobility than the rest of the protein.

3.3.2.3 Structure the HAI7 integrase

The structure of the HAI7 integrase domain (Figure 3.3.5) shows high resemblance to that of the arm type-binding domain of the HPI integrase. Positioning of helices and β-strands is identical and the RMSD of the superimposed structures (aligned taking into account the Cα atoms) is 0.81 Å. (Figure 3.3.6). The largest difference in the main chain tracing is localized in the loop between strands B3 and B4, with the HAI7 main chain bending towards helix H1. This segment shows also difference in the primary structure with small residues in HAI7 (Ser45 and Gly46) substituting large ones (Glu45 and Lys46) in HPI.
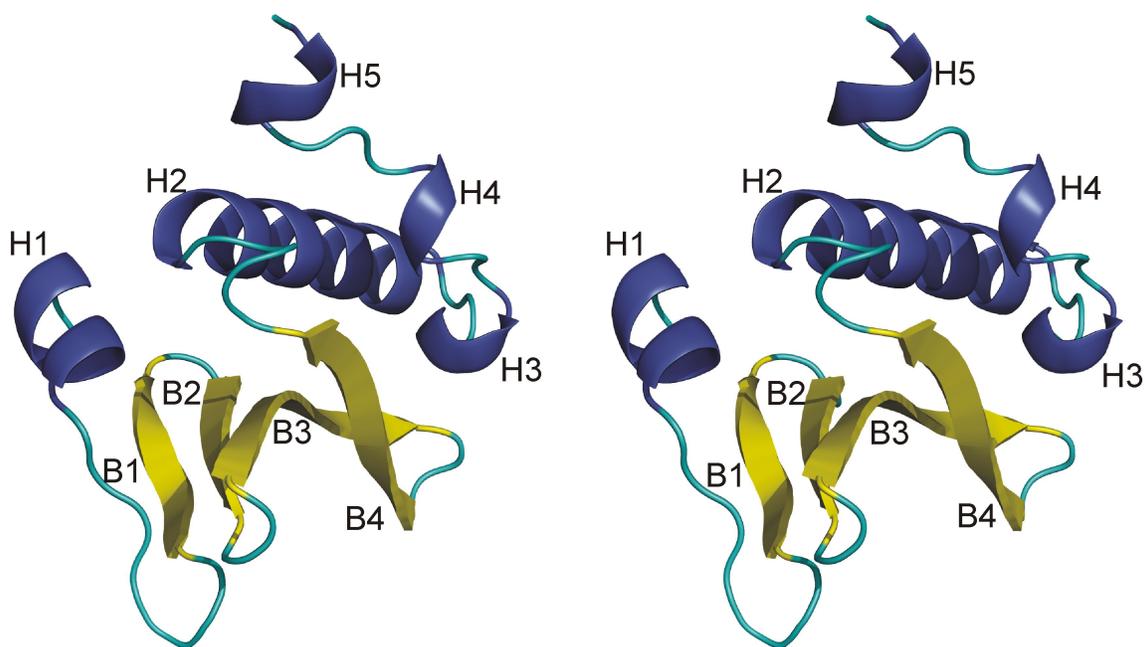
Figure 3.3.5. Crystal structure of the arm-type binding domain of the HPI integrase in ribbon representation.

The crystallized HAI7 construct is longer than that of HPI (spanning residues 1-100) and the residues located C-terminally to H2 helix are visible in the crystal structure. In disagreement with the secondary structure prediction, this region is largely unstructured with only residues 81-83, 86-88, and 92-95 forming very short helices. The electron density is however well defined, and polypeptide chain can be visualized as proceeding antiparallel to the H2 helix (Figure 3.3.7). A dense net of hydrogen bonds, formed by residues Lys73-Ser82, caps the C-terminus of the H1 helix and fixes the region in respect to the helix. The following part of the C-terminus does not form hydrogen interactions with the rest of protein, with exception of the carbonyl oxygen of Asn89 (bonding Arg67) and Asp63 (interacting with NH of Ile92).
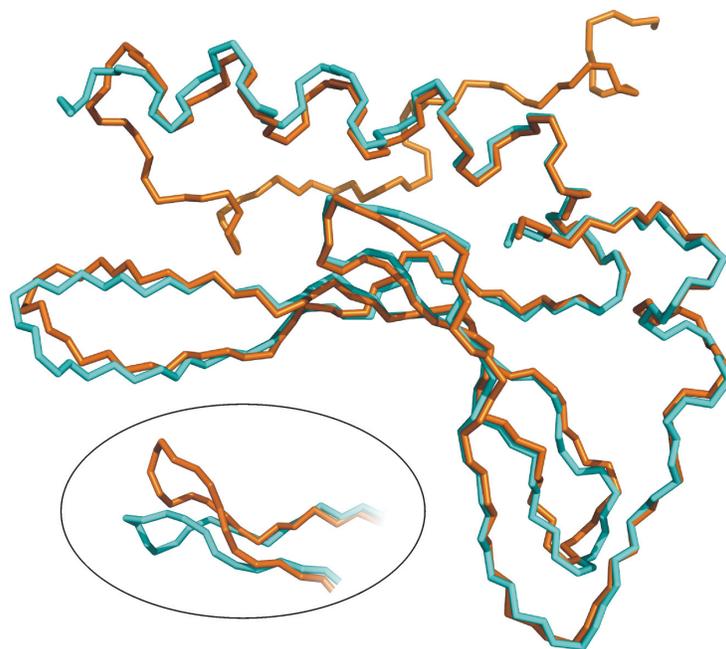
Figure 3.3.6. Alignment of the backbones of structures of the HPI (cyan) and HAI7 (orange) integrases. The insert shows magnification of the loop between strands B3 and B4.
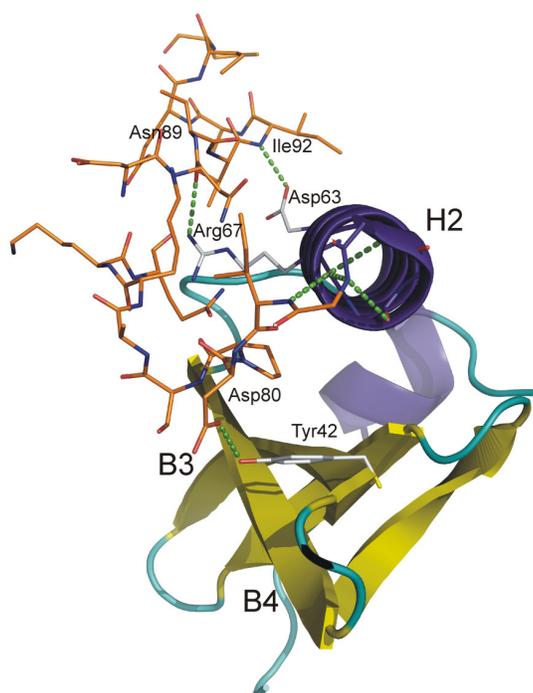
Figure 3.3.7. Interactions of the C-terminus (shown in backbone representation) with the main body of the HAI7 integrase (shown in the ribbon representation). Hydrogen bonds are colored green.

3.3.3 The complex of the HPI integrase with DNA

3.3.3.1 Oligonucleotide design and complex preparation

To elucidate the mechanism of binding to the attachment site, HPI_80 was subjected to crystallization in complex with oligonucleotides. Sequences of the DNA probes used in this study were based on the left site of the attachment P (attP_L). The oligonucleotides used in this study are listed in Table 3.3.5.

AttP_32bp covers the complete attachment region. It contains two separate binding sites (marked bold). Shorter DNA probes were designed to keep the first of the binding sites while sequentially shortening flanking regions. Probe attP_15bp_G/C contained additional bases (G or C) on the either end to introduce sticky ends (that are known to facilitate crystallization of DNA – protein complexes).

Table 3.3.5. Sequences of oligonucleotides	
Name	**Sequence**
AttP_32bp_For	CGC AAC TAT T**GG TGG TC**A TTA T**GG TGG TC**T TG
AttP_21bp_For	CGC AAC TAT T**GG TGG TC**A TTA
AttP_15bp_For	TAT T**GG TGG TC**A TTA
AttP_15bp_G _For	*G* TAT T**GG TGG TC**A TTA
AttP_12bp_For	TAT T**GG TGG TC**A
AttP_32bp_Rev	CAA **GAC CAC C**AT AAT **GAC CAC C**AA TAG TTG CG
AttP_21bp_Rev	T AAT **GAC CAC C**AA TAG TTG CG
AttP_15bp_Rev	T AAT **GAC CAC C**AA TA
AttP_15bp_C _ Rev	T AAT **GAC CAC C**AA TA*C*
AttP_12bp_Rev	T **GAC CAC C**AA TA
Binding sites are marked **bold**, sticky ends are in *italics*	

Annealing was performed by mixing two complementary oligonucleotides (dissolved in deionised water to concentration of 100 µM), heating the mixture to 95°C for 5 min and allowing it to cool to room temperature. HPI_80, purified as previously described, was added in 2:1 (for attP_32bp) or 1:1 (for remaining probes) molar ratios, incubated for 30 min in RT and concentrated afterwards.

116

3.3.3.2 Crystallization and structure determination

Complexes were subjected to screening for crystallization as previously described. For all oligonucleotides used, crystallization of the complexes resulted in crystals or crystal-like structures in multiple conditions (Figure 3.3.8). Tendency for improvement of crystal quality and reduction of crystalizable conditions could be noticed as length of oligonucleotides in complex decreased until the length of the 15 base pairs. Shorter oligonucleotides failed to produce any crystals. Introduction of the sticky ends did not improve the crystallization.

Screening of attP_32bp in complex with HPI_80 gave rise to approximately 50 crystal-like structures, none of them of diffracting qualities. attP_21bp in complex with HPI_80 formed crystals in 4 conditions (Table 3.3.1). After optimisation, condition MPD 68 produced needle-shaped crystals that did not give any diffraction pattern, when measured on a rotating anode laboratory X-ray source.

HPI_80 in complex with attP_15bp crystallized in two conditions, namely: MPD 68 and 75. Crystals from the optimized condition MPD 75 were measured on the PXII beamline of the Swiss Light Source (Villigen, Switzerland) and diffracted up to 2.5 Å. The crystals belonged to the space group $P4_12_12$ and contained two molecules per an asymmetric unit. The collected data were integrated, scaled and merged by the XDS and XSCALE programs. The Molrep program from the CCP4 suite, with the HPI_80 structure as a probe was used to determine the structure. HPI_80, having molecular mass of about 10 kDa, is of approximately equal mass to the 15-base-pair-long double-stranded oligonucleotide (9.3 kDa). Therefore, the molecular replacement was covering only 50% of the model and the remaining part of the structure (the arm-type attachment site) was build manually. The model was afterwards refined by Refmac5, and improved and revised by using XtalViev/Xfit. The

Arp/wArp program was used to add solvent atoms. The final R-factor of the complex structure is 21.9% and R-free 27.8%. Data collection and refinement statistics are summarized in Table 3.3.4.
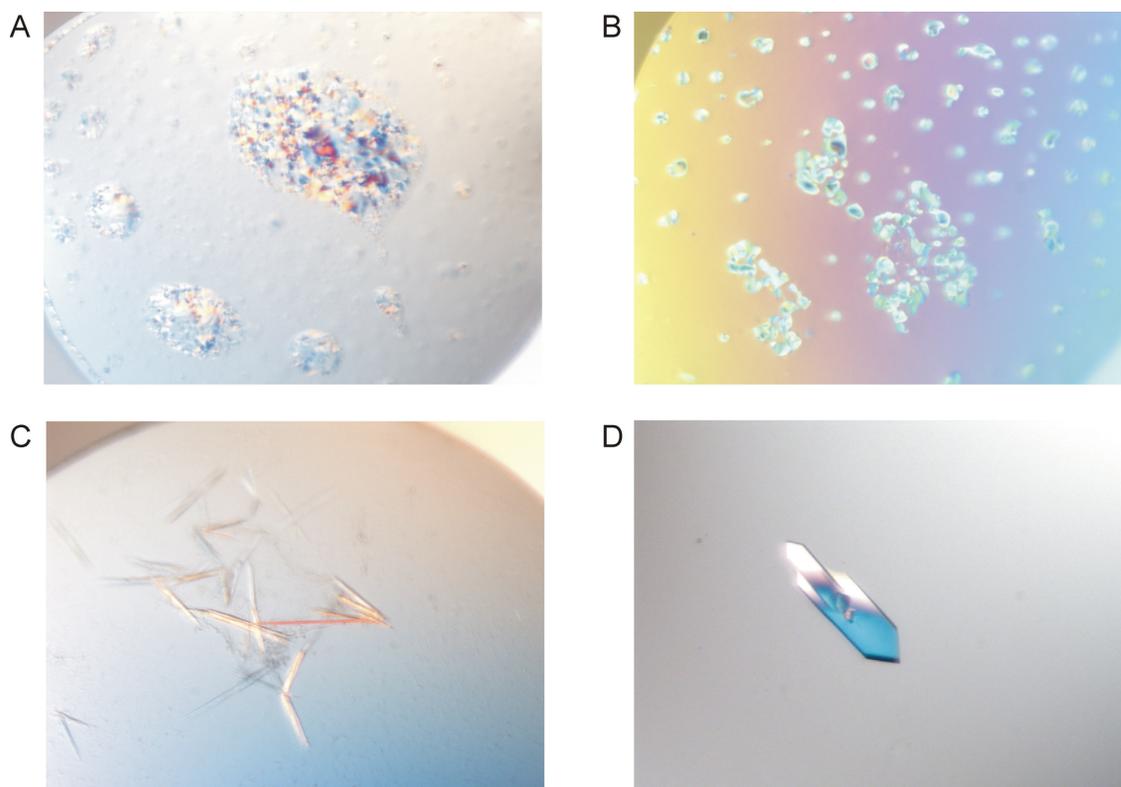

Figure 3.3.8. (A) Crystal-like structures of attP_32bp in complex with HPI_80 in Classics 71, (**B**) Crystal-like structures of attP_32bp in complex with HPI_80 in Classics 87, (**C**) Optimised crystals of attP_21bp in complex with HPI_80 in MPD 68, (**D**) X-ray quality crystal of attP_15bp in complex with HPI_80 in MPD 75.

3.3.3.3 Structure of the HPI integrase – DNA complex

Structure of the complex is presented in Figures 3.3.9 and 10. The concave surface of the β-sheet binds the major groove of the DNA, interacting directly with a streak of 9 base pairs. This interaction buries ~1100 $Å^2$ of a solvent-accessible area. The DNA molecule is bent and its major groove is widened.
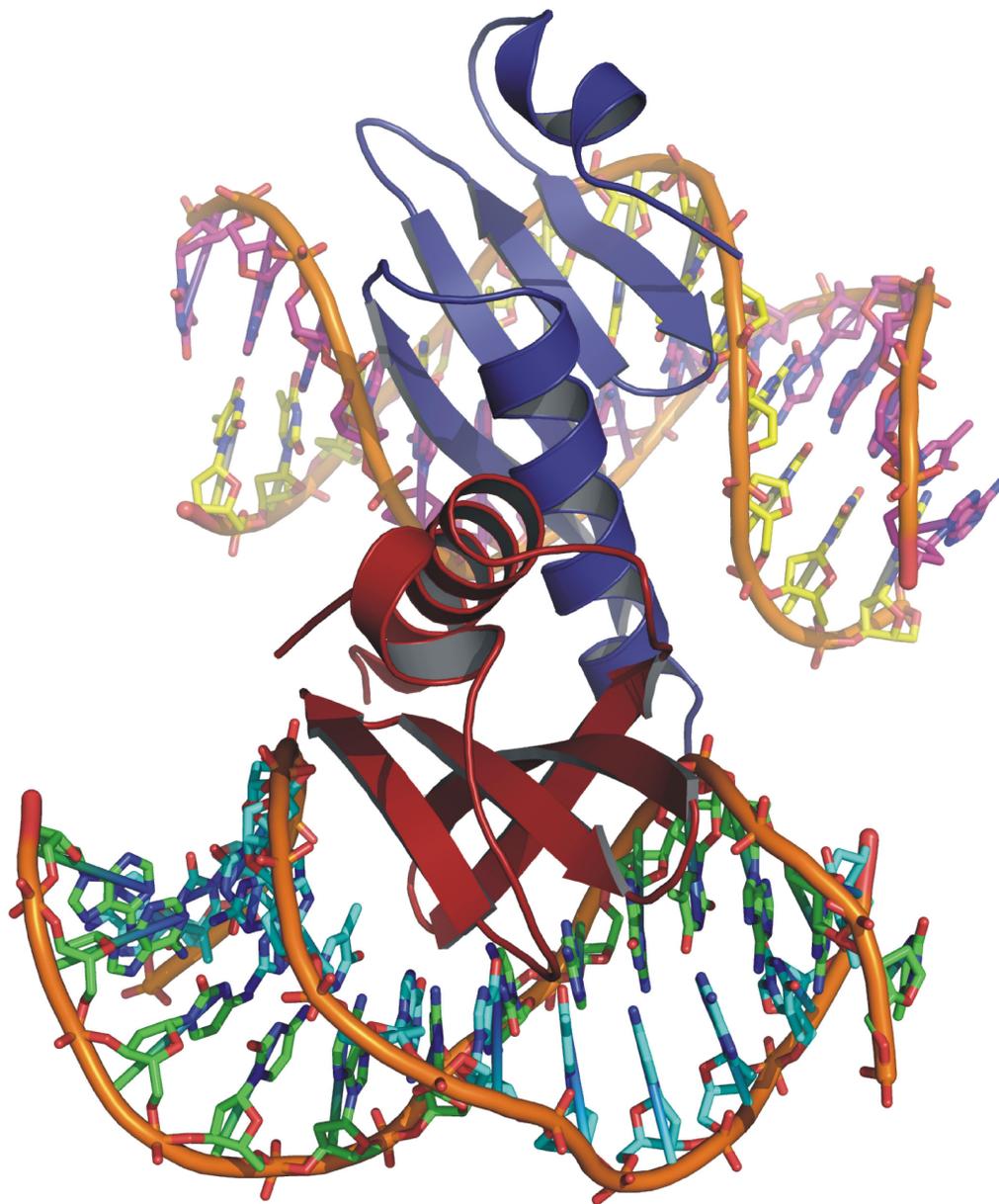
**Figure 3.3.9.** Crystal structure of the arm-type binding domain of the HPI integrase in complex with DNA.
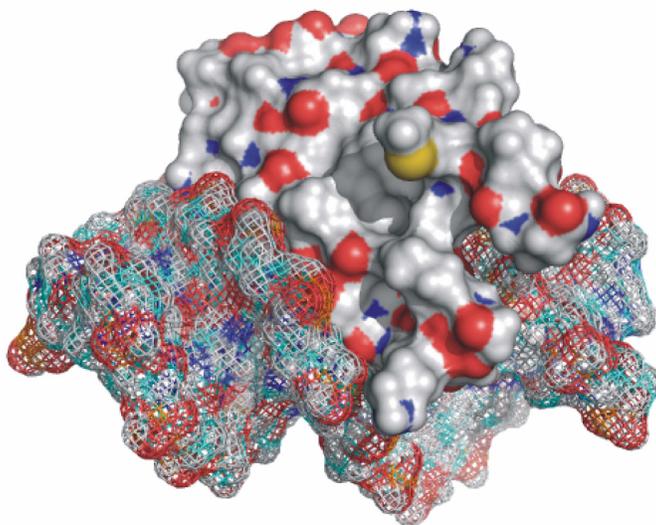
Figure 3.3.10. Solvent-accessible surfaces of the arm-type binding domain of the HPI integrase and the bound attachment site.

Two complexes present in the asymmetric unit are oriented with their H2 helices being in contact with one another. Interaction interface has the surface of ~600 Å^2. Four polar interactions were identified on this interface, namely hydrogen bonds formed by Gln66, Gln67, and Lys73 of one molecule with Asn63, Lys73, and Gln67 and Leu53, respectively, of the other (Figure 3.3.11). Hydrophobic interactions involve Gly70, Ile71, aliphatic chain of Lys73, and Met74 that closes the interface.
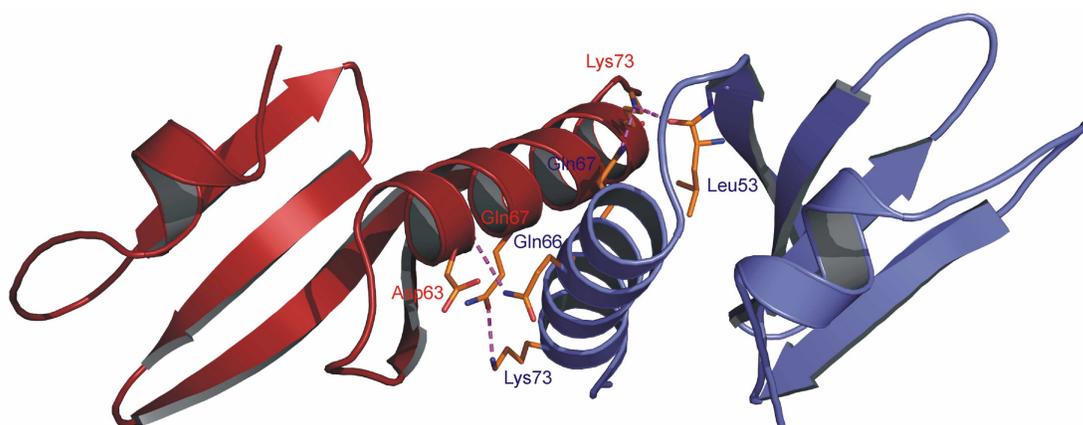

Figure 3.3.11. Hydrogen bonds of the arm type-binding domains of the HPI integrase in the asymmetric unit.

Comparison of two molecules found in the asymmetric unit shows only minor differences between them, localized in the side chains: Lys7, Lys12, Lys47, and Met74, indicating mobility of these elements. More pronounced changes can be seen in comparison with the previously solved structure of the apoprotein (Figure 3.3.12). In this case, major differences are clustered in around turn T3 (residues Arg43 – Glu48), where not only the side chains are situated differently, but the main chain tracing is shifted as well. Localization of the main chain in this region is reminiscent to that found in HAI7 although not that pronounced and directed towards H2helix, rather than away from the main body of the protein. Another region of dissimilarity is localized around residues Lys12 – Pro13, where the backbone is shifted significantly away from the main body of the protein. Turn T2 shows also small differences along with the side chains of His37 and Lys41. HAI7 is more similar to HPI bound to DNA (RMS = 0.695 Å) than to its unbound form (RMS = 0.866 Å).
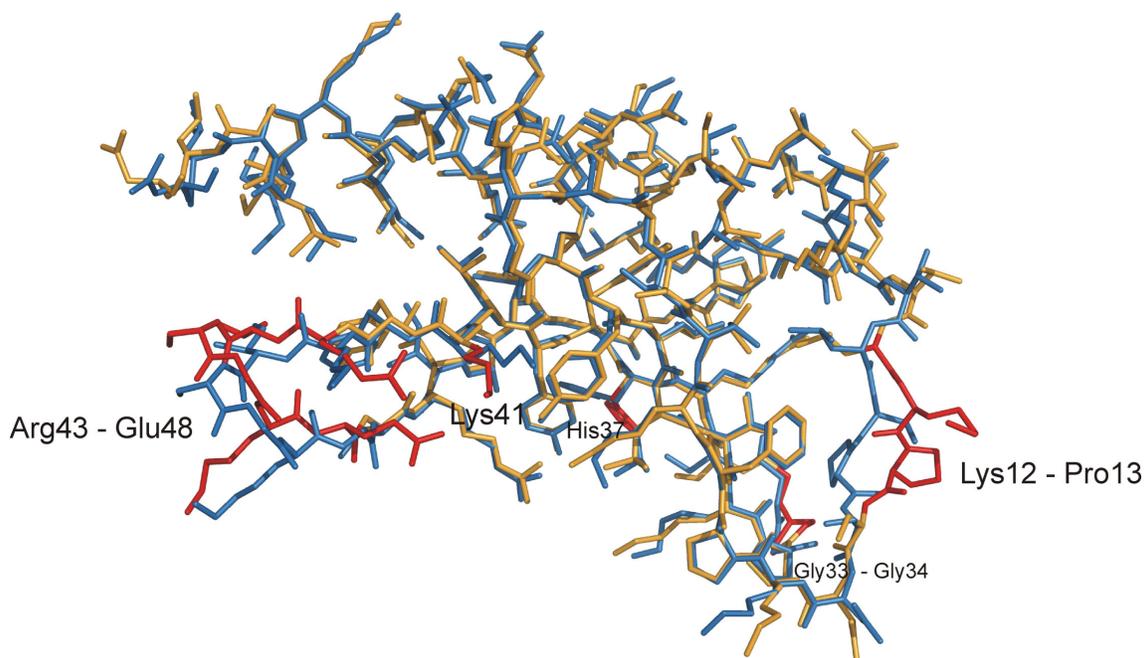
Figure 3.3.12. Comparison of the DNA-bound (orange) versus unbound (blue) arm-type binding domains of the HPI integrase. Fragments bearing the greatest differences are marked in red.

DNA binds to the protein through a dense net of polar contacts (Figure 3.3.13) Three β–strands (except for B2, which is located the furthermost from the DNA) and turn T1 participate in the interaction. Sidechains of residues Lys19, Lys41, Arg43, Glu48, and Arg50 interact with nucleotides' bases. Ser21, His24, Ser35, Tyr39, Ser49, and Ala52 bind the sugar - phosphate backbone. All specified interactions are of hydrogen bond character, except of the sole salt bridge created by Glu48 and an amide group of cytosine C^{23}.
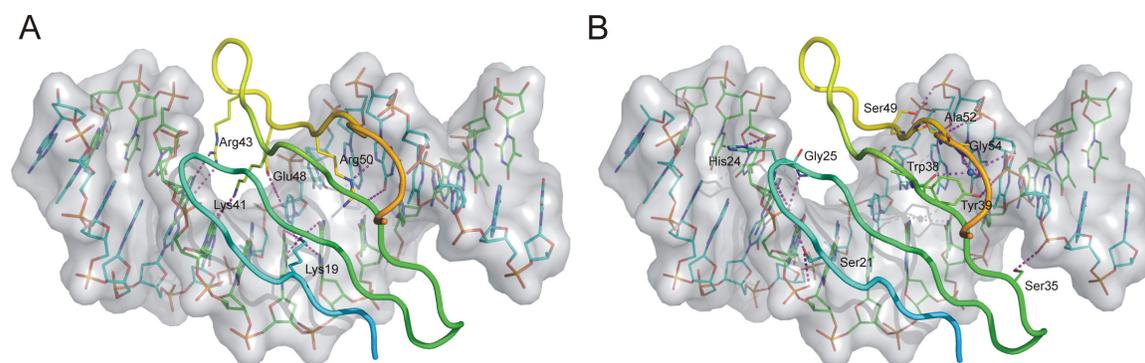

Figure 3.3.13. Protein–DNA interactions involving the bases (**A**) and the ribose – phosphate backbone (**B**) in the complex of the arm-type binding domain of the HPI integrase with its attachment site.

3.3.4. Discussion

3.3.4.1 Dimerization

In the structure of the HPI integrase in complex with DNA the protein is present as a dimer. Elaborate net of contact on the interaction interface suggests the biological, and not the crystallographical, origin of the dimer. Furthermore, careful examination of the structure of free HPI_80 reveals that the molecule from neighbouring asymmetric unit is positioned in identical way as in the dimer. Three out of four H-bonds are preserved and interaction's surface is only 10 $Å^2$ smaller. The HPI integrase appears on the gel filtration as well as in the electromobility shift assay

(EMSA) in the form of dimer, which supports the dimer hypothesis. Size exclusion chromatography as well as the electromobility shift assay (EMSA) show that the HPI-80 travels as a molecule of double the mass of the monomer; those data support the dimerization hypothesis.


'
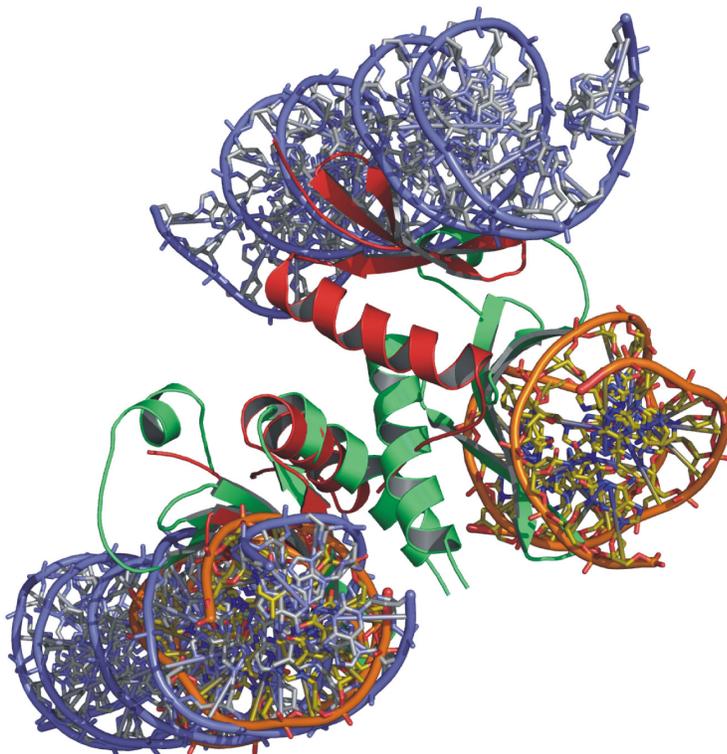
Figure 3.3.14. Alignment of the arm-type binding domains bound to their attachment sites of the lambda integrase (protein colored red, DNA colored blue) and HPI integrase (protein colored green, DNA colored yellow-orange).

The known structure of the phage lambda-derived Holliday junction (Biswas et al., 2005) features N-terminal domains oriented in the similar fashion while bound to parallel DNA strands. When compared to this assembly, the HPI_80 dimer shows significant difference in orientation of the N-terminal domains in respect to one another and, in consequence, distinguished positioning of the DNA strands (Figure

3.3.14). N-terminal domains in the phage lambda integrase are located further apart than in the HPI, with no direct interactions possible between the subunits.

3.3.4.2 The DNA binding interface

The largest observed differences between the structures of HPI_80 and HAI7_100 appear, as expected, on the DNA binding interface. Majority of the residues conserved between HPI and HAI7 that are present in the β-sheet have their side chains directed towards the core of the interface between the β-sheet and helix H2 (residues Leu28, Val30, Trp38, Leu40, and Tyr42). Three conserved residues are facing the putative DNA binding cavity, namely: Lys19, Leu29, and Arg43. Other residues pointing in the same direction differ substantially between the HPI and HAI integrases, as well as (although to the lesser extent) between HAI3 and HAI7 (Figure 3.3.15).
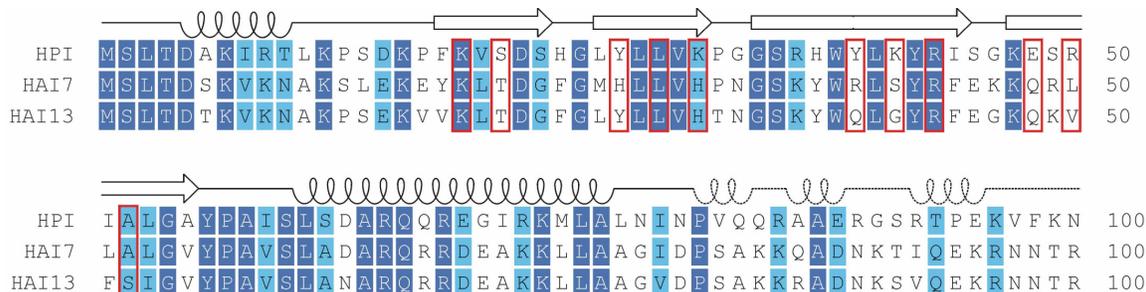

Figure 3.3.15. Alignment of arm-type binding domain of HPI integrase with and that of HAI7 and13. Identical residues are colored dark blue; similar, light blue. Strands and helices are indicated above the sequences (in case of last 20 residues, helices are presented in dotted line are taken from the structure of HAI7 integrase only). Residues pointing towards the cavity are marked red.

Comparison of putative DNA binding cavities of HAI and HPI shows several differences (Figure 3.3.15). Firstly, the proximal border of the cavity in HPI (composed of loop L1, parts of β-strands B1 and B2 and turn between them) bears a

strong positive charge introduced by Lys16, Lys19, and Lys31 (Figure 3.3.16). The situation differs in the case of HAI7, which not only lacks Lys 31 (substituted by buried histidine residues), but also presents the negative charge of Glu17 protruding from the surface of the cavity. Lys16, although present in both proteins, could not be detected in the HAI7 structure probably due to its high mobility.

The central part of the cavity in HPI (formed by middle parts of strands B3 and B4) is dominated by the positive charge of Lys41, next to the negative charge supplied by Glu48. In HAI7, the charged residues are absent - Lys 41 is substituted by serine and Glu48 by glutamine. However, Leu50 of HAI7 protrudes into the cavity approximately in the same place as the charged residues of HPI. A distal rim of the cavity in HAI7 (formed by the C-terminal end of B3, the N-terminal part of B4, and the loop between them) seems larger and bulkier than in HPI.
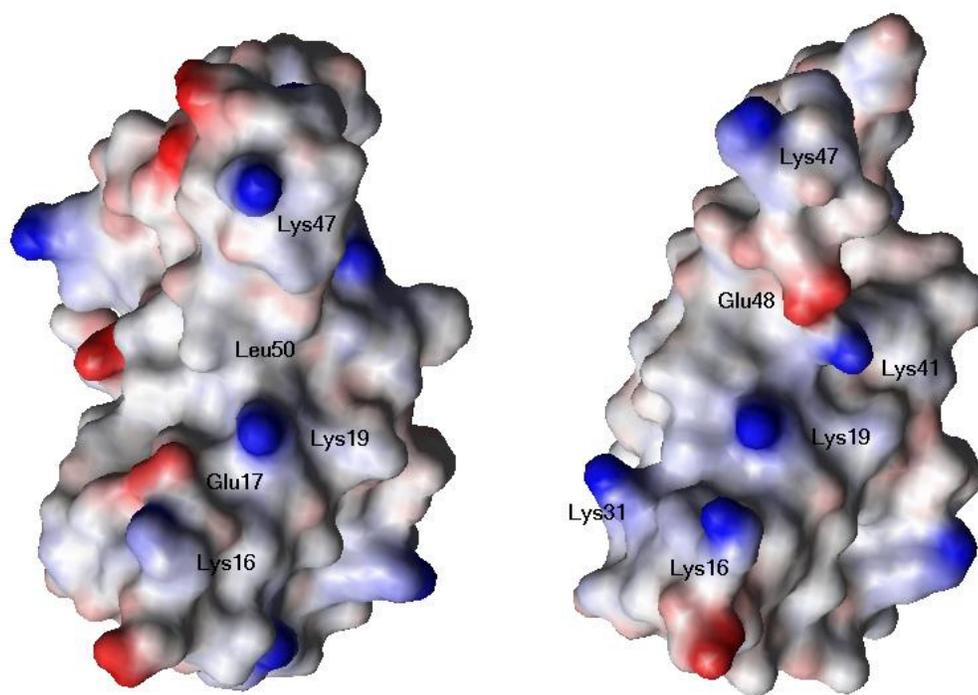
Figure 3.3.16. Comparison of DNA binding cavities of HAI7 (left) and HPI (right).

3.3.4.3 Protein – DNA interactions

Five charged residues, namely Lys19, Lys41, Arg43, Glu48 and Arg50, are responsible for DNA sequence recognition. Other residues participating in DNA binding interact with phosphate groups in the DNA backbone, contributing to stabilization of the complex (Figure 3.3.17A).
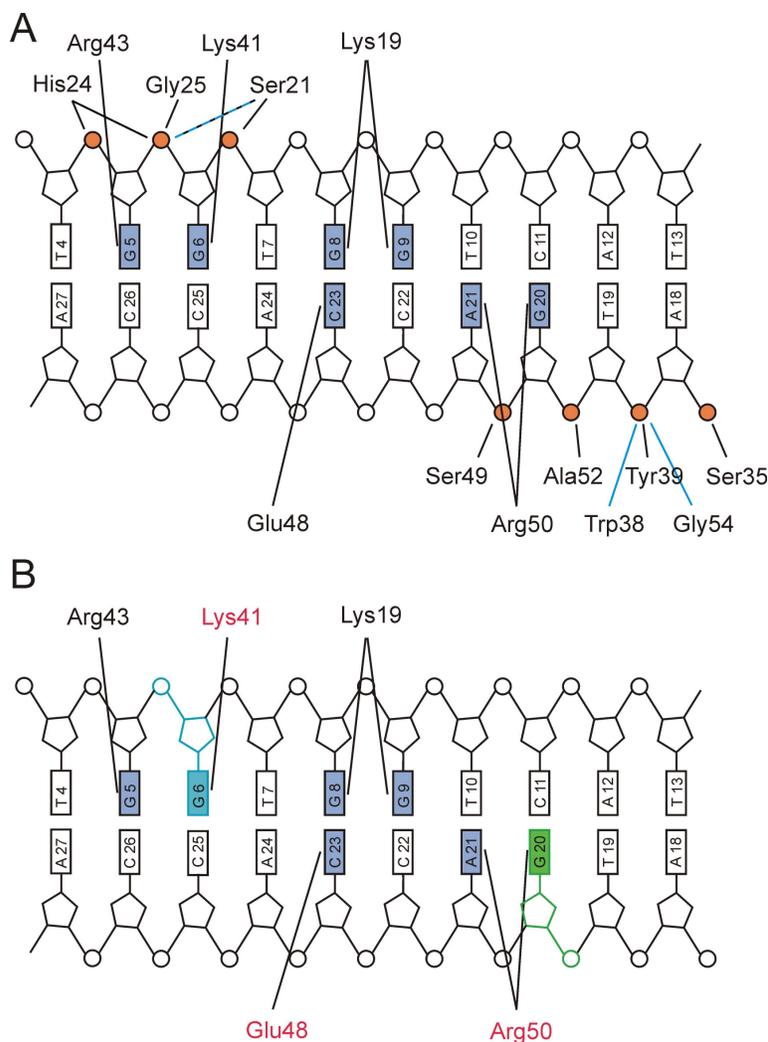
Figure 3.3.17. Schematic representation of protein – DNA interactions. Rectangles represent bases, pentagons – deoxyriboses, and circles – phosphate groups. (**A**) Phosphate groups participating in hydrogen bonds are colored orange; bases participating in hydrogen bonds are colored navy blue. Bonds created *via* interaction with a water molecule are colored blue (**B**) Residues interacting with bases are presented. Bases nonindentical between the HPI, HAI7, and HAI13 attachment sites are colored light blue and green. Protein residues nonindentical between HPI, HAI7 and HAI13 are colored red.

3.3.4.3.1 Comparison of the P1 and P2 attachment site interactions

The sequence used in crystallization trials was derived from the binding site P1 of the left attachment site and its interacting residues are identical to the neighbouring binding site, P2 as well as site P1 from the right attachment site. P2 from the right attachment site differs from the above mentioned by lacking of a centrally placed thymine, which results in the shift of the 3' part of the sequence (Figure 3.3.18).

This discrepancy in the sequence is, however, well managed by the protein interface. Lys41 and Arg43 interact with the unchanged part of the sequence. Shift of the sequence does not influence the Lys19 H-bonding. One of its partner bases, G^8 is substituted by identical one – G^9. The H-bond created by G^9 can be preserved because T^{10}, which is shifted into its place, possesses a similarly localized carbonyl oxygen. Arg50 also sustaines both of its polar contacts. Hydrogen bond with G^{20} can be preserved despite of substitution of G^{20} by T^{19} because of a similarly placed carbonyl oxygen in both bases. G^{20} subsitutes A^{21} and is able to form an interaction through its carbonyl oxygen which is localized in similar position as the amine group in adenine. Polar contact of Glu48 can be created with C^{22} instead of C^{23}.

3.3.4.3.2 Comparison of the HPI and HAI attachment site interactions

Comparison of the protein sequence of HPI and HAI integrases reveals that the two of base-recognizing residues (Lys19 and Arg43) are conserved between these integrases. When taking into account the recognized sequence of DNA (Figure 3.3.17B and 18) it can be seen that these residues interact with the bases that are common between all three integrases.

Lys41 is especially important in the context of the DNA binding specificity because it recognizes the only base (G^6, marked in blue in Figures 3.3.17B and 18)

that is different in all attachment sites of the inspected integrases. The long, charged side chain of Lys41 is substituted with serine (HAI7) or glycine (HAI13) – residues too small to reach the major groove. Spatially neighbouring residues, possibly residue 39 (arginine in HAI7 or glutamine in HAI13) or residue 27 (histidine in HAI7 or tyrosine in HAI13) may take over the Lys41 function in these integrases. Alternatively, the interaction can be dependant on a bound water molecule.

Arg50 contributes the cavity's polar contacts with three hydrogen bonds. It binds to two bases: (A^{21} and G^{20}) and to a hydroxyl group of Tyr27. Base G^{20} (marked in green in Figures 3.3.17B and 18) remains the same in the attachment site of both HAI integrases, while it is changed in HPI's. Arg50 is substituted by hydrophobic aminoacids (leucine in HAI7 and valine in HAI13). None of these residues is capable of creating polar contact due to their size and chemical character. Thus, in the HAI integrases, the recognition of the base G^{20} is probably mediated by a water molecule.

Glu48 is substituted by glutamine in both HAI7 and HAI13, which does not prevent hydrogen bond interactions.
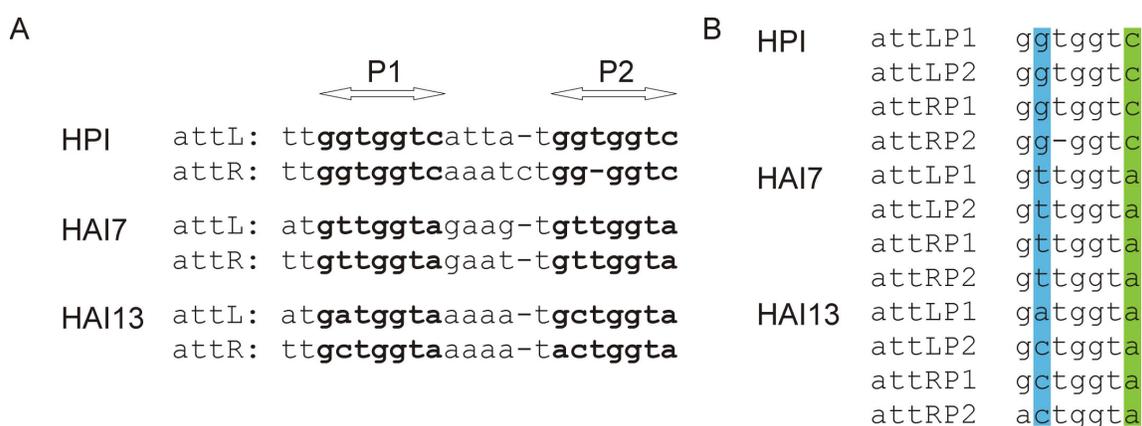
Figure 3.3.18. Left and right attachment sites of the HPI, HAI7, and HAI13 integrases. Bases that are different for each integrase are marked in blue. Bases identical between HAI7 and HAI13, while different from HPI, are marked in green.

3.3.4.4 Biological significance

Structural differences between N-terminal domains of the HPI and HAI7 integrases and of the three-stranded β-sheet DNA binding proteins support a proposal that integrases of the genomic islands form a distinct evolutionary branch of the site-specific recombinases different from that of the phage integrases. Integrases of the genomic islands are bidirectional and efficiently support both integrative and excisive recombination and do not require an additional recombination directionality factor (RDF), in contrast to the phage integrases. For example, excisive activity of the HPI integrase by its RDF is supported less than 10-fold compared to the 10^6 increase in excision of the phage lambda. Moreover, not all asn tDNA–associated islands possess RDF.

All β-strands in the structures of the arm-type binding domains of HPI and HAI7 are significantly longer than in the three-stranded arm-type binding domains, this thus enables mimicking the shape of the major groove of a longer DNA fragment. The HPI – recognized fragment comprises of nine nucleotides, one base longer than the attachment site of the lambda integrase. In the lambda integrase, however, only six nucleotides are recognized through the interaction of the β-sheets with the major groove, while remaining two are bound by the N-terminal helix inserting into the minor groove (Fadeev et al., 2009). The HPI-type integrases encompass the whole attachment-binding site by means of its extended β-sheet. The function of the N-terminal helix remains unknown, but it can be speculated that it plays a role in the RDF-independent excision.

4 SUMMARY

The subject of this thesis was to structurally characterize three proteins: nicastrin (a subunit of the γ-secretase), the cyanobacterial biliprotein lyases, and the *Enterobacteriaceae*-derived genomic islands' integrases.

γ-secretase is proposed to function as a "membrane proteasome", clearing the membrane of processed proteins. It has arelaxed substrate specificity because shedding of the extracellular domain in single-pass transmembrane proteins is a sufficient requirement for their cleavage. Two of γ-secretase substrates, Notch and Alzheimer precursor protein (APP), have gained special attention due to their biological significance. The former is the initial element in the Notch signalling cascade, which controls cell-cell communication during differentiation processes. The intramembrane cleavage of the latter results in production of the neurotoxic β-amyloid – the causative agent of the Alzheimer disease. Because of this link with the disease that is the leading cause of dementia in the aging society, γ-secretase makes a potential drug target for therapy. However, due to the multitude of its targets, the detailed structure of γ-secretase must be uncovered to eliminate the possibility of blocking the fundamental, along with the pathogenic processes. In this study, the substrate-recognizing subunit of γ-secretase, nicastrin, was examined with the aim to gain insight into its structure and to open the possibilities of screening for active compounds.

The second part of the thesis focused on the enzymes which catalyze the attachment of the chromophore to the photoactive biliproteins – structural components of the light-harvesting antennas in cyanobacteria. The chain tetrapyrrol bilin prosthetic group is linked via a tioether bond to a cysteine residue of the protein.

Numerous types of bilins, as well as multiple apoproteins, entail the need for an assembly of lyases necessary to synthesize the complete phycobilisome. The main focus of this study was the determination of the structural mechanisms of the catalysis. Additionally, various biochemical and biophysical techniques, including NMR spectroscopy, gel filtration, as well as reconstitution assays, were employed to investigate the interactions in the phycocyanobilin system from *Mastigocladus laminosus, Synechococcus* sp. PCC7002, and *Nostoc* sp. PCC7120.

Structural investigations on the integrases from the genomic islands of *Enterobacteriaceae* were the subject of the third part of this thesis. Horizontally acquired genomic elements bear some level of independency from their bacterial host, which is achieved through coding of elements of the site-specific recombination system (integrases and the attachment sites). In this thesis I present the structures of the arm-type binding domains of the integrases derived from *Yersinia pestis* and *Erwinia carotovora*. These structures can be classified as a new type of folds in the arm-type binding domains, due to the presence of additional N-terminal elements: a helix and a β-strand. The structure of the complex of the *Yersinia pestis* integrase with DNA gives insight into the base recognition mode. Additionally, the dimerization interface of the HPI integrase could be determined. This interface can play a role in the functionality of the whole integrase.

5 Zusammenfassung

In dieser Arbeit wurden drei Proteine strukturell charakterisiert: Nicastrin (eine Untereinheit der γ-Sekretase), cyanobakterielle Biliprotein Lyasen und Integrasen aus den genomischen Inseln von *Enterobacteriaceaen*.

Die γ-Sekretase ist ein „Membranproteasom", welches prozessierte Proteine abbaut. Es besitzt eine niedrige Substratsspezifität. Ein „Singlepass"-Transmembranprotein, dessen extrazelluläre Domäne entfernt worden ist, ist bereits ein ausreichendes Abbausignal. Zwei der γ-Sekretase Substrate, Notch und Alzheimer precursor protein (APP), haben durch ihre biologische Signifikanz an Bedeutung gewonnen. Ersteres ist das Ausgangselement der Notch-Signalkaskade, welches die Zell-Zell Kommunikation während der Differenzierungsprozesse kontrolliert. Die intramembrane Spaltung des Letzteren führt zu dem neurotoxischen Produkt ß-amyloid, welches verantwortlich für die Alzheimer Krankheit ist. Aufgrund des Zusammenhangs mit dieser maßgeblichen Ursache für Demenz in der alternden Gesellschaft, stellt die γ-Sekretase ein wichtiges Ziel für die Suche eines potenziellen Arzneimittels dar. Um jedoch eine gezielte Wirkung auf die pathogenen Prozesse zu erzielen, ohne die fundamentalen Aufgaben des Proteins zu beeinflussen, ist es notwendig die Struktur der γ-Sekretase aufzuklären. Ziel dieser Arbeit war es die Substraterkennungsuntereinheit der γ -Sekretase strukturell zu untersuchen, um eine gezielte Suche nach aktiven Substraten zu ermöglichen.

Der zweite Teil der Arbeit befasst sich mit Enzymen, welche die Anhaftung von Chromophoren an die photoaktiven Biliproteine (strukturelle Komponente des Lichtsammelkomplexes) der Cyanobakterien katalysiert. Die prostethische Gruppe des Tetrapyrolbilins ist über eine Thioetherbindung an den Cysteinrest des Proteins geknüpft. Das vollständige Phycobilisom besteht aus verschiedenen Arten von

Bilinen, sowie mehreren Apoproteinen. Der Schwerpunkt dieser Arbeit lag in der strukturellen Aufklärung des Katalysemechanismus. Zudem wurden verschiedene biochemische und biophysikalische Techniken, unter anderem NMR-Spektroskopie, Gelfiltration und Rekonstruktionsexperimente angewandt, um die Interaktion des Phycocyanobilin-Systems aus *Mastigocladus laminosus, Synechococcus sp. PCC7002* und *Nostoc sp. PCC7120* zu untersuchen.

Im dritten Teil dieser Arbeit wurden Integrasen aus den genomischen Inseln von Enterobacteriaceaen untersucht. Genomische Elemente die über einen horizontalen Gentransfer erworben wurden, sind zu einem gewissen Grad unabhängig von deren Wirt, was durch das ortspezifische Rekombinasesystem (Integrasen und Bindungsstellen) erreicht wird. In dieser Arbeit präsentiere ich die Strukturen der arm-ähnlichen Bindungsdomänen der Integrasen aus *Yersinia pestis* und E*rwinia carotovora*. Diese Strukturen können aufgrund der zusätzlichen N-terminalen α-Helix und eines β-Stanges als eine neue Art arm-ähnlicher Bindedomänen klassifiziert werden. Die Struktur des Integrase/DNA-Komplexes aus *Yersinia pestis* gibt Einsicht in die grundlegende Erkennungsmethode. Zusätzlich konnte die Dimerisierungsschnittstelle der HPI-Integrase, welche eine funktionelle Bedeutung in der gesamten Integrase haben kann, bestimmt werden.

6 BIBLIOGRAPHY

Adir, N. (2005). Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. *Photosynth.Res., 85,* 15-32.

Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998) A novel mode of DNA recognition by a b-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO,* 17, 5484–5496.

Anderson, L. K. and Toole, C. M. (1998). A model for early events in the assembly pathway of cyanobacterial phycobilisomes. *Mol.Microbiol., 30,* 467-474.

Annaert, W. G., Esselens, C., Baert, V., Boeve, C., Snellings, G., Cupers, P. et al. (2001). Interaction with telencephalin and the amyloid precursor protein predicts a ring structure for presenilins. *Neuron, 32,* 579-589.

Antonenka, U., Nolting, C., Heesemann, J., and Rakin, A. (2005). Horizontal transfer of Yersinia high-pathogenicity island by the conjugative RP4 attB target-presenting shuttle plasmid. *Mol.Microbiol., 57,* 727-734.

Bach, S., de, A. A., and Carniel, E. (2000). The Yersinia high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiol.Lett., 183,* 289-294.

Beel, A. J. and Sanders, C. R. (2008). Substrate specificity of gamma-secretase and other intramembrane proteases. *Cell Mol.Life Sci., 65,* 1311-1334.

Bell, K. S., Sebaihia, M., Pritchard, L., Holden, M. T., Hyman, L. J., Holeva, M. C. et al. (2004). Genome sequence of the enterobacterial phytopathogen Erwinia carotovora subsp. atroseptica and characterization of virulence factors. *Proc.Natl.Acad.Sci.U.S.A, 101,* 11105-11110.

Benedek, O. and Schubert, S. (2007). Mobility of the Yersinia High-Pathogenicity Island (HPI): transfer mechanisms of pathogenicity islands (PAIS) revisited (a review). *Acta Microbiol.Immunol.Hung., 54,* 89-105.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature, 435,* 1059-1066.

Bohm, S., Endres, S., Scheer, H., and Zhao, K. H. (2007). Biliprotein chromophore attachment: chaperone-like function of the PecE subunit of alpha-phycoerythrocyanin lyase. *J.Biol.Chem., 282,* 25357-25366.

Brown, M. S. and Goldstein, J. L. (1999). A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc.Natl.Acad.Sci.U.S.A, 96,* 11041-11048.

Buchrieser, C., Brosch, R., Bach, S., Guiyoule, A., and Carniel, E. (1998). The high-pathogenicity island of Yersinia pseudotuberculosis can be inserted into any of the three chromosomal asn tRNA genes. *Mol.Microbiol., 30,* 965-978.

Carniel, E. (2001). The Yersinia high-pathogenicity island: an iron-uptake island. *Microbes.Infect., 3,* 561-569.

Chavez-Gutierrez, L., Tolia, A., Maes, E., Li, T., Wong, P. C., and De, S. B. (2008). Glu(332) in the Nicastrin ectodomain is essential for gamma-secretase complex maturation but not for its activity. *J.Biol.Chem., 283,* 20096-20105.

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol.Cell, 6,* 885-897.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr., 50*, 760–763.

Connolly, K. M., Wojciak, J. M., and Clubb, R. T. (1998). Site-specific DNA binding using a variation of the double stranded RNA binding motif. *Nat. Struct. Biol., 5*, 546-550.

Crystal, A. S., Morais, V. A., Pierson, T. C., Pijak, D. S., Carlin, D., Lee, V. M. et al. (2003). Membrane topology of gamma-secretase component PEN-2. *J.Biol.Chem., 278,* 20117-20123.

De La Fortelle, and E., Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement in MIR and MAD methods. *Meth Enzym., 276*, 472-494.

Fadeev, E. A., Sam, M. D., and Clubb, R. T. (2009). NMR structure of the amino-terminal domain of the lambda integrase protein in complex with DNA: immobilization of a flexible tail facilitates beta-sheet recognition of the major groove. *J.Mol.Biol., 388,* 682-690.

Fagan, R., Swindells, M., Overington, J., and Weir, M. (2001). Nicastrin, a presenilin-interacting protein, contains an aminopeptidase/transferrin receptor superfamily domain. *Trends Biochem.Sci., 26,* 213-214.

Freeman, M. (2008). Rhomboid proteases and their biological functions. *Annu.Rev.Genet., 42,* 191-210.

Gal-Mor, O. and Finlay, B. B. (2006). Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell Microbiol., 8,* 1707-1719.

Glazer, A. N. (1982). Phycobilisomes: structure and dynamics. *Annu.Rev.Microbiol., 36,* 173-198.

Grainge, I. and Jayaram, M. (1999). The integrase family of recombinase: organization and function of the active site. *Mol.Microbiol., 33,* 449-456.

Grindley, N. D., Whiteson, K. L., and Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu.Rev.Biochem., 75,* 567-605.

Grossman, A. R., Schaefer, M. R., Chiang, G. G., and Collier, J. L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol.Rev., 57,* 725-749.

Groth, A. C. and Calos, M. P. (2004). Phage integrases: biology and applications. *J.Mol.Biol., 335,* 667-678.

Herreman, A., Van, G. G., Bentahir, M., Nyabi, O., Craessaerts, K., Mueller, U. et al. (2003). gamma-Secretase activity requires the presenilin-dependent trafficking of nicastrin through the Golgi apparatus but not its complex glycosylation. *J.Cell Sci., 116,* 1127-1136.

Jiang, T., Zhang, J. P., Chang, W. R., and Liang, D. C. (2001). Crystal structure of R-phycocyanin and possible energy transfer pathways in the phycobilisome. *Biophys.J., 81,* 1171-1179.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol.Rev., 33,* 376-393.

Jung, L. J., Chan, C. F., and Glazer, A. N. (1995). Candidate genes for the phycoerythrocyanin alpha subunit lyase. Biochemical analysis of pecE and pecF interposon mutants. *J.Biol.Chem., 270,* 12877-12884.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst., 26*, 795-800.

Kaether, C., Capell, A., Edbauer, D., Winkler, E., Novak, B., Steiner, H. et al. (2004). The presenilin C-terminus is required for ER-retention, nicastrin-binding and gamma-secretase activity. *EMBO J., 23,* 4738-4748.

Kahn, K., Mazel, D., Houmard, J., Tandeau de, M. N., and Schaefer, M. R. (1997). A role for cpeYZ in cyanobacterial phycoerythrin biosynthesis. *J.Bacteriol., 179,* 998-1006.

Kim, S. H. and Sisodia, S. S. (2005). A sequence within the first transmembrane domain of PEN-2 is critical for PEN-2-mediated endoproteolysis of presenilin 1. *J.Biol.Chem., 280,* 1992-2001.

Kopan, R. and Goate, A. (2000). A common enzyme connects notch signaling and Alzheimer's disease. *Genes Dev., 14,* 2799-2806.

Kopan, R. and Ilagan, M. X. (2009). The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell, 137,* 216-233.

Kornilova, A. Y., Bihel, F., Das, C., and Wolfe, M. S. (2005). The initial substrate-binding site of gamma-secretase is located on presenilin near the active site. *Proc.Natl.Acad.Sci.U.S.A, 102,* 3230-3235.

Kurokawa, Y., Yanagi, H., and Yura, T. (2001). Overproduction of bacterial protein disulfide isomerase (DsbC) and its modulator (DsbD) markedly enhances periplasmic production of human nerve growth factor in Escherichia coli. *J.Biol.Chem., 276,* 14393-14399.

Lai, E. C. (2002). Notch cleavage: Nicastrin helps Presenilin make the final cut. *Curr.Biol., 12,* R200-R202.

Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models. *Acta Cryst.*, *D49*, 129-47.

Lazarov, V. K., Fraering, P. C., Ye, W., Wolfe, M. S., Selkoe, D. J., and Li, H. (2006). Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores. *Proc.Natl.Acad.Sci.U.S.A, 103,* 6889-6894.

Liu, C. Y. and Kaufman, R. J. (2003). The unfolded protein response. *J.Cell Sci., 116,* 1861-1862.

MacColl, R. (1998). Cyanobacterial phycobilisomes. *J.Struct.Biol., 124,* 311-334.

MacColl, R. (2004). Allophycocyanin and energy transfer. *Biochim.Biophys.Acta, 1657,* 73-81.

Martin, S. S., Pulido, E., Chu, V. C., Lechner, T. S., and Baldwin, E. P. (2002). The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP Holliday junction complex. *J.Mol.Biol., 319,* 107-127.

Masters, C. L., Cappai, R., Barnham, K. J., and Villemagne, V. L. (2006). Molecular mechanisms for Alzheimer's disease: implications for neuroimaging and therapeutics. *J.Neurochem., 97,* 1700-1725.

McRee, D. E. (1999) XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J Struc Biol., 125*, 156-65.

Munter, L. M., Voigt, P., Harmeier, A., Kaden, D., Gottschalk, K. E., Weise, C. et al. (2007). GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of Abeta42. *EMBO J., 26,* 1702-1712.

Niimura, M., Isoo, N., Takasugi, N., Tsuruoka, M., Ui-Tei, K., Saigo, K. et al. (2005). Aph-1 contributes to the stabilization and trafficking of the gamma-secretase complex through mechanisms involving intermolecular and intramolecular interactions. *J.Biol.Chem., 280,* 12967-12975.

Pardossi-Piquard, R., Yang, S. P., Kanemoto, S., Gu, Y., Chen, F., Bohm, C. et al. (2009). APH1 polar transmembrane residues regulate the assembly and activity of presenilin complexes. *J.Biol.Chem., 284,* 16298-16307.

Periz, G. and Fortini, M. E. (2004). Functional reconstitution of gamma-secretase through coordinated expression of presenilin, nicastrin, Aph-1, and Pen-2. *J.Neurosci.Res., 77,* 309-322.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct Bio.,* 6, 458-63.

Radman-Livaja, M., Biswas, T., Ellenberger, T., Landy, A., and Aihara, H. (2006). DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Curr.Opin.Struct.Biol., 16,* 42-50.

Rakin, A., Noelting, C., Schropp, P., and Heesemann, J. (2001). Integrative module of the high-pathogenicity island of Yersinia. *Mol.Microbiol., 39,* 407-415.

Rehm, T., Huber, R., and Holak, T. A. (2002). Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure, 10*, 1613-1618.

Reuter, W., Wiegand, G., Huber, R., and Than, M. E. (1999). Structural analysis at 2.2 A of orthorhombic crystals presents the asymmetry of the allophycocyanin-linker complex, AP.LC7.8, from phycobilisomes of Mastigocladus laminosus. *Proc.Natl.Acad.Sci.U.S.A, 96,* 1363-1368.

Ross, W. and Landy, A. (1982). Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. *Proc.Natl.Acad.Sci.U.S.A, 79,* 7724-7728.

Sadowski, P. (1986). Site-specific recombinases: changing partners and doing the twist. *J.Bacteriol., 165,* 341-347.

Sato, C., Morohashi, Y., Tomita, T., and Iwatsubo, T. (2006). Structure of the catalytic pore of gamma-secretase probed by the accessibility of substituted cysteines. *J.Neurosci., 26,* 12081-12088.

Saunee, N. A., Williams, S. R., Bryant, D. A., and Schluchter, W. M. (2008). Biogenesis of phycobiliproteins: II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to CYS-82 OF beta-phycocyanin and CYS-81 of allophycocyanin subunits in Synechococcus sp. PCC 7002. *J.Biol.Chem., 283,* 7513-7522.

Scheer, H. and Zhao, K. H. (2008). Biliprotein maturation: the chromophore attachment. *Mol.Microbiol., 68,* 263-276.

Schmidt, M., Krasselt, A., and Reuter, W. (2006). Local protein flexibility as a prerequisite for reversible chromophore isomerization in alpha-phycoerythrocyanin. *Biochim. Biophys. Acta, 1764,* 55-62.

Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr.,* 58,1772-1779.

Schubert, S., Dufke, S., Sorsa, J., and Heesemann, J. (2004). A novel integrative and conjugative element (ICE) of Escherichia coli: the putative progenitor of the Yersinia high-pathogenicity island. *Mol.Microbiol., 51,* 837-848.

Shah, S., Lee, S. F., Tabuchi, K., Hao, Y. H., Yu, C., LaPlant, Q. et al. (2005). Nicastrin functions as a gamma-secretase-substrate receptor. *Cell, 122,* 435-447.

Shaikh, A. C. and Sadowski, P. D. (1997). The Cre recombinase cleaves the lox site in trans. *J.Biol.Chem., 272,* 5695-5702.

Shen, G., Schluchter, W. M., and Bryant, D. A. (2008). Biogenesis of phycobiliproteins: I. cpcS-I and cpcU mutants of the cyanobacterium Synechococcus sp. PCC 7002 define a heterodimeric phyococyanobilin lyase specific for beta-phycocyanin and allophycocyanin subunits. *J.Biol.Chem., 283,* 7503-7512.

Sibanda, B. L., Blundell, T. L., and Thornton, J. M. (1989). Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol., 206*, 759-777.

Six, C., Thomas, J. C., Thion, L., Lemoine, Y., Zal, F., and Partensky, F. (2005). Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium Synechococcus sp. strain WH8102. *J.Bacteriol., 187,* 1685-1694.

Storf, M., Parbel, A., Meyer, M., Strohmann, B., Scheer, H., Deng, M. G. et al. (2001). Chromophore attachment to biliproteins: specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-cys-alpha 84-phycoviolobilin chromophore of phycoerythrocyanin. *Biochemistry, 40,* 12444-12456.

Titball, R. W., Hill, J., Lawton, D. G., and Brown, K. A. (2003). Yersinia pestis and plague. *Biochem.Soc.Trans., 31,* 104-107.

Tolia, A., Chavez-Gutierrez, L., and De, S. B. (2006). Contribution of presenilin transmembrane domains 6 and 7 to a water-containing cavity in the gamma-secretase complex. *J.Biol.Chem., 281,* 27633-27642.

Tolia, A. and De, S. B. (2009). Structure and function of gamma-secretase. *Semin.Cell Dev.Biol., 20,* 211-218.

Tolia, A., Horre, K., and De, S. B. (2008). Transmembrane domain 9 of presenilin determines the dynamic conformation of the catalytic site of gamma-secretase. *J.Biol.Chem., 283,* 19793-19803.

Tomita, T. (2008). At the frontline of Alzheimer's disease treatment: gamma-secretase inhibitor/modulator mechanism. *Naunyn Schmiedebergs Arch.Pharmacol., 377,* 295-300.

Toth, I. K., Pritchard, L., and Birch, P. R. (2006). Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu.Rev.Phytopathol., 44,* 305-336.

Tu, J. M., Zhou, M., Haessner, R., Ploscher, M., Eichacker, L., Scheer, H. et al. (2009). Toward a Mechanism for Biliprotein Lyases: Revisiting Nucleophilic Addition to Phycocyanobilin. *J.Am.Chem.Soc..*

Van Duyne, G. D. (2001). A structural view of cre-loxp site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct., 30*, 87-104.

Watanabe, N., Tomita, T., Sato, C., Kitamura, T., Morohashi, Y., and Iwatsubo, T. (2005). Pen-2 is incorporated into the gamma-secretase complex through binding to transmembrane domain 4 of presenilin 1. *J.Biol.Chem., 280,* 41967-41975.

Wiegand, G., Parbel, A., Seifert, M. H., Holak, T. A., and Reuter, W. (2002). Purification, crystallization, NMR spectroscopy and biochemical analyses of alpha-phycoerythrocyanin peptides. *Eur.J.Biochem., 269,* 5046-5055.

Wojciak, J. M., Connolly, K. M., and Clubb, R. T. (1999). NMR structure of the Tn916 integrase-DNA complex. *Nat.Struct.Biol., 6,* 366-373.

Wojciak, J. M., Sarkar, D., Landy, A., and Clubb, R. T. (2002). Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. *Proc. Natl. Acad. Sci. U. S. A, 99,* 3434-3439.

Xia, W. and Wolfe, M. S. (2003). Intramembrane proteolysis by presenilin and presenilin-like proteases. *J.Cell Sci., 116,* 2839-2844.

Yamasaki, A., Eimer, S., Okochi, M., Smialowska, A., Kaether, C., Baumeister, R. et al. (2006). The GxGD motif of presenilin contributes to catalytic function and substrate identification of gamma-secretase. *J.Neurosci., 26,* 3821-3828.

Yu, G., Nishimura, M., Arawaka, S., Levitan, D., Zhang, L., Tandon, A. et al. (2000). Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature, 407,* 48-54.

Zhao, K. H., Deng, M. G., Zheng, M., Zhou, M., Parbel, A., Storf, M. et al. (2000). Novel activity of a phycobiliprotein lyase: both the attachment of phycocyanobilin and the isomerization to phycoviolobilin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. *FEBS Lett., 469,* 9-13.

Zhao, K. H., Su, P., Bohm, S., Song, B., Zhou, M., Bubenzer, C. et al. (2005a). Reconstitution of phycobilisome core-membrane linker, LCM, by autocatalytic chromophore binding to ApcE. *Biochim.Biophys.Acta, 1706,* 81-87.

Zhao, K. H., Su, P., Li, J., Tu, J. M., Zhou, M., Bubenzer, C. et al. (2006). Chromophore attachment to phycobiliprotein beta-subunits: phycocyanobilin:cysteine-beta84 phycobiliprotein lyase activity of CpeS-like protein from Anabaena Sp. PCC7120. *J.Biol.Chem., 281,* 8573-8581.

Zhao, K. H., Su, P., Tu, J. M., Wang, X., Liu, H., Ploscher, M. et al. (2007a). Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. *Proc.Natl.Acad.Sci.U.S.A, 104,* 14300-14305.

Zhao, K. H., Wu, D., Wang, L., Zhou, M., Storf, M., Bubenzer, C. et al. (2002). Characterization of phycoviolobilin phycoerythrocyanin-alpha 84-cystein-lyase-(isomerizing) from Mastigocladus laminosus. *Eur.J.Biochem., 269,* 4542-4550.

Zhao, K. H., Wu, D., Zhou, M., Zhang, L., Bohm, S., Bubenzer, C. et al. (2005b). Amino acid residues associated with enzymatic activities of the isomerizing phycoviolobilin-lyase PecE/F. *Biochemistry, 44,* 8126-8137.

Zhao, K. H., Zhang, J., Tu, J. M., Bohm, S., Ploscher, M., Eichacker, L. et al. (2007b). Lyase activities of CpcS- and CpcT-like proteins from Nostoc PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. *J.Biol.Chem., 282,* 34093-34103.

Zhou, J., Gasparich, G. E., Stirewalt, V. L., de, L. R., and Bryant, D. A. (1992). The cpcE and cpcF genes of Synechococcus sp. PCC 7002. Construction and phenotypic characterization of interposon mutants. *J.Biol.Chem., 267,* 16138-16145.

Abbreviations

1D	one-dimensional
2D	two-dimensional
Å	Ångstrøm (10-10 m)
aa	amino acid
Aβ	β-amyloid
AICD	intracellular domain of APP
AP	alkaline phosphatase
APC	allophycocyanin
Aph-1	anterior pharynx defective–1
APS	ammonium peroxodisulfate
APP	alzheimer precursor protein
attB	host attachment site
attL	left attachment site
attP	phage attachment site
attR	right attachment site
BACE1	β-site APP-cleaving enzyme 1
bp	base pair
BSA	bovine serum albumin
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
cDNA	complimentary DNA
CIP	calf intestinal alkaline phosphatase
CPC	phycocyanin
CTF	C-terminal fragment
Da	Dalton (g mol-1)
DAP	DYIGS and the peptidase domain
DAPT	N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DR	direct repeat

DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic amid
EM	electron microscopy
ER	endoplasmic reticulum
g	gravity (9.81 m s-2)
GEI	genomic island
GSH	reduced glutathione
GSSG	oxidized glutathione
HAI	horizontally acquired island
HD	hydrophobic domain
HEPES	N-(2-hydroxyethyl)piperazine-N`-(2-ethanesulfonic acid)
HGT	horizontal gene transfer
HPI	high pathogenicity island
HSQC	heteronuclear single quantum coherence
Hz	Hertz
ICD	intracellular domain
IMAC	immobilized metal affinity chromatography
I-CLIPs	intramembrane-cleaving proteases
IPTG	isopropyl-b-thiogalactopyranoside
LB	Luria-Broth medium
LIC	ligation-independant cloning
MM	minimal medium
MPD	2-methyl-2,4-pentanediol
MW	molecular weight
NCT	nicastrin
NECD	Notch extracellular domain
NEXT	Notch extracellular truncation
NICD	Notch intracellular domain
NiNTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NTF	N-terminal fragment
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis

PAI	pathogenecity island
PBS	phosphate-buffered saline
PCB	phycocyanobilin
PE	phycoerythrin
PEB	phycoerythrobilin
PEC	phycoerythrocyanin
PEG	polyethylene glycol
PEN-2	presenilin enhancer–2
PS	presenilin
ppm	parts per million
PUB	phycourobilin
PVB	phycoviolobilin
PVDF	polyvinylidene fluoride
RIP	regulated intramembrane proteolysis
S2P	site 2 protease
SDS	sodium dodecyl sulphate
SPP	signal peptide peptidase
SREBP	sterol-regulatory-element binding protein
TAE	Tris-acetate-EDTA buffer
TB	terrific broth
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylendiamine
TMD	transmembrane domain
Tris	tris(hydroxymethyl)aminomethane
YNB	yeast nitrogen base
YPD	yeast extract peptone dextrose
YPDS	yeast extract peptone dextrose sorbitol

Amino acids and nucleotides are abbreviated according to either one or three letter IUPAC code.